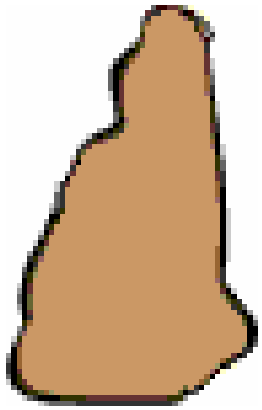
New Hampshire Department of Education

The Nations' Report Card
New Hampshire
Department of Education
2003 NAEP State Report
Grade 8 Reading



FORWARD

Although this report was put together in final form by New Hampshire Department of Education staff there are a number of other significant contributors who made its outcome possible.

First, we acknowledge the many schools' students and staff who gave of their time and energy to participate in the 2003 State National Assessment of Educational Progress (NAEP). As the New Hampshire sample they allowed an estimate of what grade four and grade eight students in our state and the nation know and can do in mathematics and reading. Without them of course there would be no data; nothing to report. The last time New Hampshire had data of this type was in 1998 so the 2003 State NAEP assessment was a significant event.

Equally as important is the work done by the National Center of Education Statistics and its contractors who systematically gathered, scored, and organized the results in usable tables and graphs. This work made the monumental task of ferreting out recognizable results manageable, providing valuable opportunities for analysis. We are in debt as well to the wonderful and helpful people at the NAEP State Service Center. they provided excellent training and support on a continual basis to assure the highest level of success in all the state NAEP endeavors.

Special recognition as well is given to the designers of the State Report Generator (SRG) that allowed customized state report generation from the voluminous data gathered. The selection and filtering mechanisms made possible the generation of this and other New Hampshire NAEP reports in a timely fashion. Nancy Mead and her colleagues at the Educational Testing Service gave us a superb product to utilize.

Finally, recognition and thanks is given to the many persons in the Bureau of Accountability at the New Hampshire Department of Education who provided guidance and expertise in shaping the final report products. A special "Thank You" is set aside for Carol Angowski whose creative and technical skill was essential in producing this and a number of New Hampshire NAEP-related published documents.

Contact Information
Testing Director: Tim Kurtz
NAEP Coordinator: David Gebhardt
NH State Dept. of Education

Table of Contents

Eighth Grade Reading Report

Overall Student, School/District Characteristics 2002-2003

Student Characteristics

Number enrolled: **207,671**
Percent in Title I schools: **48.7%**
With Individualized Education Programs (IEP): **13.9%**
Percent in limited-English proficiency programs: **1.57%**
Percent eligible for free/reduced lunch: **16.39%**

Racial/Ethnic Background

White: **94.4%**
Black: **1.5%**
Hispanic: **2.2%**
Asian/Pacific Islander: **1.6%**
American Indian/Alaskan Native: **0.3%**

School/District Characteristics

Number of SAUs: **84**
Number of school districts: **176**
Number of schools: **466**
Number of charter schools: **N/A**
Per-pupil expenditures: **$7,233**[1]
Pupil/teacher ratio: **13.5**
Number of FTE teachers: **14,478**

'--' : data unavailable Source: Common Core of Data, *2002-2003* school year
[1] Common Core of Data, *2002-2003* school year

The Nations' Report Card
READING 2003

New Hampshire
Grade 8
Public Schools

NEW HAMPSHIRE NAEP STATE REPORT

KEY FINDINGS

For grade 8:

- The average reading scale score for students in New Hampshire was 271.
- New Hampshire's average score (271) was higher than that of the nation's public schools (261).
- Students' average scores in New Hampshire were higher than those in 40 jurisdictions, and not significantly different from those in 12 jurisdictions.
- The percentage of students in New Hampshire who performed at or above the *Proficient* level was 40 percent.
- In New Hampshire, the percentage of students who performed at or above *Proficient* was higher than that for the nation's public schools (30 percent).

This report provides selected results from the National Assessment of Educational Progress (NAEP) for New Hampshire's public-school students at grade 8. Since 1992, reading has been assessed in five different years at the state level (at grade 4 in 1992 and 1994, and at both grades 4 and 8 in 1998, 2002, and 2003). In 2003, 53 jurisdictions participated: the 50 states, District of Columbia, Department of Defense Domestic Dependent Elementary and Secondary Schools, and Department of Defense Dependents Schools (Overseas). New Hampshire participated and met the criteria for reporting public-school results at grade 4 in 1992, 1994, 1998, and at both grades 4 and 8 in 2003.

NAEP is a project of the National Center for Education Statistics (NCES). For more information about the assessment, see *The Nation's Report Card, Reading Highlights 2003* or *The Nation's Report Card: Reading 2003*, which will be available in 2004. The full set of results is available in an interactive database on the NAEP web site (http://nces.ed.gov/nationsreportcard/). Released test questions, scoring guides, and question-level performance data are also available on the web site.

The U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) has provided software that generated user-selectable data, statistical significance test result statements, and technical descriptions of the NAEP assessments for this report. Content may be added or edited by states or other jurisdictions. This document, therefore, is not an official publication of the National Center for Education Statistics.

Introduction

What Was Assessed?

The content for each NAEP assessment is determined by the National Assessment Governing Board (NAGB). The development process for reading required the active participation of teachers, curriculum specialists, subject-matter specialists, local school administrators, parents, and members of the general public. The objectives for each NAEP assessment are described in a "framework," a document that delineates the important content and process areas to be measured, as well as the types of questions to be included on the assessment. The reading framework is available on the NAGB web site (http://www.nagb.org/pubs/read_fw_03.pdf).

The reading framework for the 1992 and 1994 reading assessments also guided the 1998, 2000 (national grade 4 only), 2002, and 2003 assessments. This framework was developed under the auspices of the Council of Chief State School Officers (CCSSO) and directed by NAGB. In 2002, the framework was updated to provide more explicit detail regarding the assessment design. In the process, some of the terms used to describe elements of the reading assessment were altered slightly. It should be noted, however, that these alterations do not represent a change in the content or design of the NAEP reading assessment.

The framework is founded on a body of research from the field of education that defines reading as an interactive and constructive process involving the reader, the text, and the context of the reading experience. Reading involves the development of an understanding of text, thinking about the text in different ways, and using a variety of text types for different purposes.

Recognizing that readers vary their approach to reading different texts, the framework specifies the assessment of reading in three contexts: reading for literary experience, reading to gain information, and reading to perform a task. Each context for reading is associated with a range of different types of texts that are included in the NAEP reading assessment. All three contexts for reading are assessed at grades 8 and 12, but reading to perform a task is not assessed at grade 4.

As readers attempt to develop an understanding of a text, they focus on general topics or themes, interpret and integrate ideas, make connections to background knowledge and experiences, and examine the content and structure of the text. The framework accounts for these different approaches to understanding text by specifying four "aspects of reading" that represent the types of comprehension questions asked of students. All four aspects of reading are assessed at all three grades within each context for reading. The reading framework specifies the percentage distribution of questions by grade level for each of the contexts for and aspects of reading.

The assessment contains reading materials that were drawn from sources commonly available to students both in and out of the school environment. These authentic materials were considered to be representative of students' typical reading experiences. Each student in the state assessment was asked to complete two 25-minute sections, each consisting of a reading passage and associated comprehension questions. A combination of multiple-choice and constructed-response questions was used to assess students' understanding of the passages. Released NAEP reading passages and questions, along with student performance data by state, are available on the NAEP web site (http://nces.ed.gov/nationsreportcard/itmrls/).

Who Was Assessed?

In 2003, 53 jurisdictions participated in NAEP: the 50 states, District of Columbia, Department of Defense Domestic Elementary and Secondary Schools, and Department of Defense Dependents Schools (Overseas). The target sample for each state or other jurisdiction was approximately 100 schools at a grade and approximately 3,000 students for each subject at a grade, except in small or sparsely populated jurisdictions. The sample of schools and students was chosen in a two-stage sampling process. First, the sample of schools was selected by probability sampling methods. Then, within the participating schools, random samples of students were chosen. Beginning in 2002, the national sample was obtained by aggregating the samples from each state. The national results include the results from the states, weighted appropriately to represent the U.S. student population. Only public schools, however, are included in the state reports. The overall participation rates for schools and students must meet guidelines established by the National Center for Education Statistics (NCES) and the National Assessment Governing Board (NAGB) in order for assessment results to be reported publicly. Data are not reported to the public for a state or jurisdiction that participates but does not meet minimum participation guidelines (see http://nces.ed.gov/nationsreportcard/about/participrates.asp). Participation rates for the 2003 reading assessment are available at the NAEP web site (http://nces.ed.gov/nationsreportcard/reading/sampledesign.asp).

How Is Student Reading Performance Reported?

The results of student performance on the NAEP assessments are reported for various groups of students (e.g., fourth-grade female students or students who took the assessment in different years). NAEP does not produce scores for individual students or report scores for schools. Nor are data produced for school districts, except that some large urban districts voluntarily participated in the assessment on a trial basis and were sampled as states were sampled. Reading performance for groups of students is reported in two ways: 1) average scale scores and 2) achievement levels.

Scale Scores: Student performance is reported as an average score based on the NAEP reading scale, which ranges from 0 to 500 and is linked to the corresponding scales in 1992, 1994, 1998, 2000, and 2002. Subscales were created to report performance on each of the contexts for reading defined in the NAEP reading framework. An overall composite scale was developed by weighting each of the reading subscales for the grade (two at grade 4 and three at grade 8) based on its relative importance in the framework. This composite scale is the metric used to present the average scale scores and selected percentiles used in NAEP reports.

Achievement Levels: Student reading performance is also reported in terms of three achievement levels—*Basic, Proficient*, and *Advanced*. Results based on achievement levels are expressed in terms of the percentage of students who attained each level. The three achievement levels are defined as follows:

- *Basic:* This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.
- *Proficient:* This level represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.
- *Advanced:* This level signifies superior performance.

The achievement levels are performance standards adopted by the National Assessment Governing Board (NAGB) as part of its statutory responsibilities mandated by Congress. The levels represent collective judgments of what students should know and be able to do for each grade tested. They are based on recommendations made by broadly representative panels of classroom teachers, education specialists, and members of the general public. As provided by law, the National Center for Education Statistics (NCES), upon review of congressionally mandated evaluations of NAEP, has determined that the achievement levels are to be used on a trial basis until it is determined that the achievement levels are "reasonable, valid, and informative to the public."[1] However, both NCES and NAGB believe these performance standards are useful for understanding trends in student achievement. They have been widely used by national and state officials as a common yardstick for academic performance. The reading achievement-level descriptions are summarized in figure 1.

Students With Disabilities (SD) and/or Limited-English-Proficient (LEP) Students

The results displayed in this report and official publications of NAEP 2003 results are based on representative samples that include students with disabilities (SD) and limited-English-proficient students (LEP). Some of these students were assessed using accommodations that allowed them to participate. In state NAEP reading assessments prior to 1998, no testing accommodations or adaptations were permitted for special-needs students in these samples. However, research carried out by NAEP showed that the results for such accommodated students could be combined with the results for nonaccommodated students without compromising the validity of the NAEP scales in trend comparisons. Therefore, the special-needs students who typically received accommodations in their classroom testing, and who required these accommodations to participate, also received them in the NAEP assessment, provided the accommodations did not change the nature of what was tested.

In 1998, NAEP used a split sample of schools—one sample in which accommodations were permitted for special-needs students who normally received them and another sample in which accommodations were not permitted. Therefore, there are two different sets of results displayed for 1998. Results for the assessment years where accommodations were not permitted in state NAEP assessments (1992, 1994, and 1998) are reported in the same tables as the results where accommodations were permitted (1998, 2002, and 2003). The results labeled "Accommodations not permitted" are based on the same procedures as previously reported data. The results labeled "Accommodations permitted" for 1998 are based on the new procedures.

Statistical comparisons are made between the results across years, regardless of accommodation conditions, because NAEP's statistical studies showed that these comparisons could be made and the results remain valid. For 1998, when accommodations were permitted for one sample and not for another sample, comparisons to both samples are available in tables and in the data tool (http://nces.ed.gov/nationsreportcard/naepdata/). In the text of this report, comparisons to the 1998 results are discussed only for the sample for which accommodations were permitted.

Cautions in Interpreting Results

The averages and percentages in this report have a standard error—a range of up to a few points above or below the score—which takes into account potential score fluctuation due to sampling error and measurement error. Statistical tests that factor in these standard errors are used to determine whether the differences between average scores or percentages are significant. All differences were tested for statistical significance at the 0.05 level. NAEP sample sizes have increased since 2002 compared to previous years, resulting in smaller standard errors. As a consequence, smaller differences are detected as statistically significant than in previous assessments.

In this report, statistically significant differences are referred to as "significant differences" or "significantly different." Significant differences between 2003 and prior assessments are marked with a notation (*) in the tables. Any differences in scores within a year or across years that are mentioned in the text as "higher," "lower," "greater," or "smaller" are statistically significant.

Estimates based on small subgroups are likely to have large standard errors. Consequently some seemingly large differences may not be statistically significant. The reader is cautioned to rely on reported differences in the tables and/or text, which are statistically significant, rather than on the apparent magnitude of any difference. Readers are also cautioned against interpreting NAEP results causally. Inferences related to subgroup performance, for example, should take into account the many socioeconomic and educational factors that may affect student performance.

---

1. No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2001).

The Nation's Report Card 2003 State Assessment

Descriptions of NAEP reading achievement levels, grade 8

Basic Level (243)	Eighth-grade students performing at the *Basic* level should demonstrate a literal understanding of what they read and be able to make some interpretations. When reading text appropriate to eighth grade, they should be able to identify specific aspects of the text that reflect the overall meaning, extend the ideas in the text by making simple inferences, recognize and relate interpretations and connections among ideas in the text to personal experience, and draw conclusions based on the text.

For example, when reading **literary** text, *Basic*-level eighth graders should be able to identify themes and make inferences and logical predictions about aspects such as plot and characters.
When reading **informational** text, they should be able to identify the main idea and the author's purpose. They should make inferences and draw conclusions supported by information in the text. They should recognize the relationships among the facts, ideas, events, and concepts of the text (e.g., cause and effect and chronological order).
When reading **practical** text, they should be able to identify the main purpose and make predictions about the relatively obvious outcomes of procedures in the text.

Proficient Level (281)	Eighth-grade students performing at the *Proficient* level should be able to show an overall understanding of the text, including inferential as well as literal information. When reading text appropriate to eighth grade, they should be able to extend the ideas in the text by making clear inferences from it, by drawing conclusions, and by making connections to their own experiences—including other reading experiences. *Proficient* eighth graders should be able to identify some of the devices authors use in composing text.

For example, when reading **literary** text, students at the *Proficient* level should be able to give details and examples to support themes that they identify. They should be able to use implied as well as explicit information in articulating themes; to interpret the actions, behaviors, and motives of characters; and to identify the use of literary devices such as personification and foreshadowing.
When reading **informational** text, they should be able to summarize the text using explicit and implied information and support conclusions with inferences based on the text.
When reading **practical** text, *Proficient*-level students should be able to describe its purpose and support their views with examples and details. They should be able to judge the importance of certain steps and procedures.

Advanced Level (323)	Eighth-grade students performing at the *Advanced* level should be able to describe the more abstract themes and ideas of the overall text. When reading text appropriate to eighth grade, they should be able to analyze both meaning and form and support their analyses explicitly with examples from the text, and they should be able to extend text information by relating it to their experiences and to world events. At this level, student responses should be thorough, thoughtful, and extensive.

For example, when reading **literary** text, *Advanced*-level eighth graders should be able to make complex, abstract summaries and theme statements. They should be able to describe the interactions of various literary elements (i.e., setting, plot, characters, and theme) and explain how the use of literary devices affects both the meaning of the text and their response to the author's style. They should be able critically to analyze and evaluate the composition of the text.
When reading **informational** text, they should be able to analyze the author's purpose and point of view. They should be able to use cultural and historical background information to develop perspectives on the text and be able to apply text information to broad issues and world situations.
When reading **practical** text, *Advanced*-level students should be able to synthesize information that will guide their performance, apply text information to new situations, and critique the usefulness of the form and content.

SOURCE: National Assessment Governing Board. (2002). *Reading Framework for the 2003 National Assessment of Educational Progress*. Washington, DC: Author.

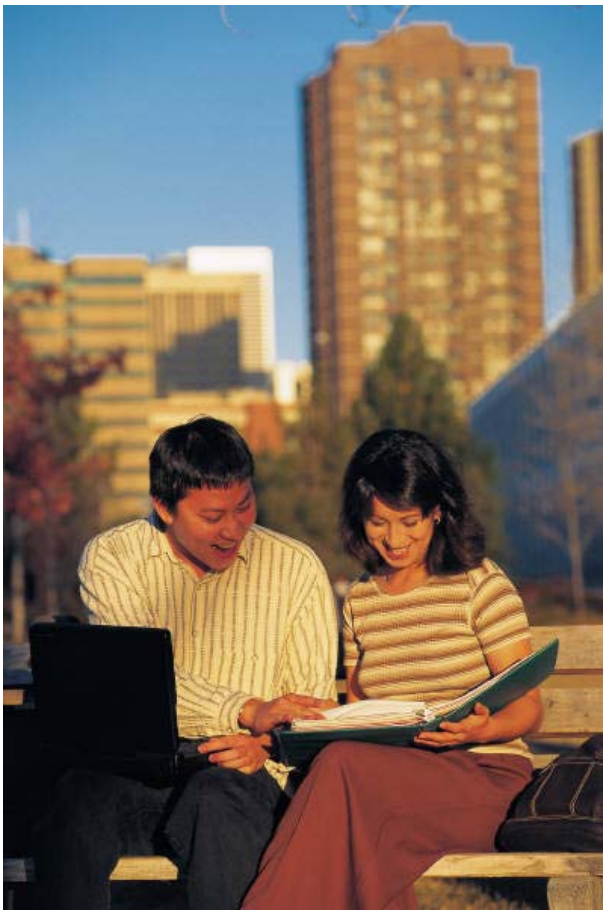
NAEP Reading 2003 Overall Scale Score and Achievement-Level Results for Public School Students

Overall Scale Score Results

In this section student performance is reported as an average score based on the NAEP reading scale, which ranges from 0 to 500. Scores on this scale are comparable from 1992 through 2003.

Prior to 1998, testing accommodations were not provided for students with special needs in state reading assessments. In 1998 only, results were reported for two samples of students: one in which accommodations were permitted and one in which accommodations were not pemitted. Subsequent assessment results were based on the more inclusive samples. In the text of this report, comparisons to 1998 results refer only to the sample in which accommodations were permitted.

Table 1 shows the overall performance results of grade 8 public school students in New Hampshire and the nation. The first column of results presents the average score on the NAEP reading scale. The subsequent columns show the score at selected percentiles. The percentile indicates the percentage of students who performed below the score for that percentile. For example, 10 percent of the students had scores that were lower than the score shown for the 10th percentile.



Grade 8 Scale Score Results

- In 2003, the average scale score for students in New Hampshire was 271. This was higher than that of students across the nation (261).

TABLE 1

The Nation's Report Card 2003 State Assessment

Average reading scale scores and selected percentiles, grade 8 public schools: 2003

	Average Scale Score	Scale score distribution				
		10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
Accommodations permitted						
2003 New Hampshire	271 (0.9)	227 (2.5)	251 (1.3)	273 (1.4)	293 (1.0)	310 (1.8)
Nation (Public)	261 (0.2)	215 (0.5)	240 (0.3)	264 (0.3)	286 (0.3)	304 (0.3)

NOTE: The NAEP reading scale ranges from 0 to 500. The standard errors of the statistics in the table appear in parentheses. All differences were tested for statistical significance at the 0.05 level using unrounded numbers. Performance comparisons may be affected by differences in exclusion rates for students with disabilities and limited-English-proficient students in the NAEP samples and changes in sample sizes. NAEP sample sizes have increased since 2002 compared to previous years, resulting in smaller detectable differences than in previous assessments.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.

Overall Achievement-Level Results

In this section student performance is reported as the percentage of students performing relative to standards set by the National Assessment Governing Board (NAGB). These performance standards for what students should know and be able to do were based on the recommendations of broadly representative panels of educators and members of the public. In 1998 only, results were obtained for student samples for which accommodations were permitted and were not permitted. However, in the text of this report, comparisons to 1998 results refer only to the sample in which accommodations were permitted.

Table 2 presents the percentage of students at grade 8 who performed below *Basic*, at or above *Basic*, at or above *Proficient*, and at the *Advanced* level. Because the percentages are cumulative from *Basic* to *Proficient* to *Advanced*, they sum to more than 100 percent. Only the percentage of students performing at or above *Basic* (which includes the students at *Proficient* and *Advanced*) plus the students below *Basic* will sum to 100 percent (except for rounding).

Grade 8 Achievement-Level Results

- In 2003, the percentage of New Hampshire's students who performed at or above the *Proficient* level was 40 percent. This was greater than the percentage of the nation's public school students who performed at or above *Proficient* (30 percent).

TABLE 2

The Nation's Report Card 2003 State Assessment

Percentage of students at or above each reading achievement level, grade 8 public schools: 2003

		At or above		
	Below *Basic*	At or above *Basic*	*Proficient*	*Advanced*
Accommodations permitted				
2003 New Hampshire	19 (1.2)	81 (1.2)	40 (1.5)	4 (0.4)
Nation (Public)	28 (0.3)	72 (0.3)	30 (0.3)	3 (0.1)

NOTE: The standard errors of the statistics in the table appear in parentheses. Achievement levels correspond to the following points on the NAEP reading scale: below *Basic*, 242 or lower; *Basic*, 243-280; *Proficient*, 281-322; and *Advanced*, 323 and above. All differences were tested for statistical significance at the 0.05 level using unrounded numbers. Details may not sum to totals due to rounding. Performance comparisons may be affected by differences in exclusion rates for students with disabilities and limited-English-proficient students in the NAEP samples and changes in sample sizes. NAEP sample sizes have increased since 2002 compared to previous years, resulting in smaller detectable differences than in previous assessments.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.

Comparisons Between New Hampshire and Other Participating States and Jurisdictions

In 2003, 53 jurisdictions participated in the reading assessment. These include the 50 states, the District of Columbia and the two groups of Department of Defense Education Activity (DoDEA) schools: Domestic Dependent Elementary and Secondary Schools (DDESS) and Department of Defense Dependents Schools (DoDDS).

Comparisons by Average Scale Scores

Figure 2 compares New Hampshire's 2003 overall reading scale scores at grade 8 with those of all other participating states and jurisdictions. The different shadings indicate whether a state's or jurisdiction's average scale score was found to be higher than, lower than, or not significantly different from that of New Hampshire in the NAEP 2003 reading assessment.

Grade 8 Scale Score Comparisons Results

- Students' scale scores in New Hampshire were higher than those in 40 jurisdictions, and not significantly different from those in 12 jurisdictions.

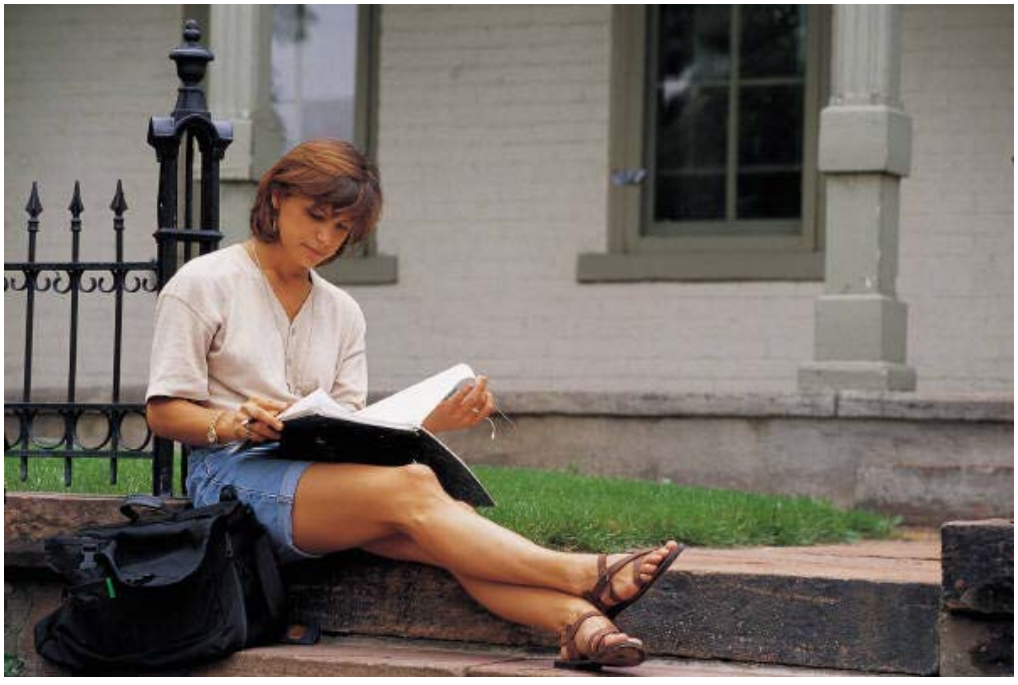
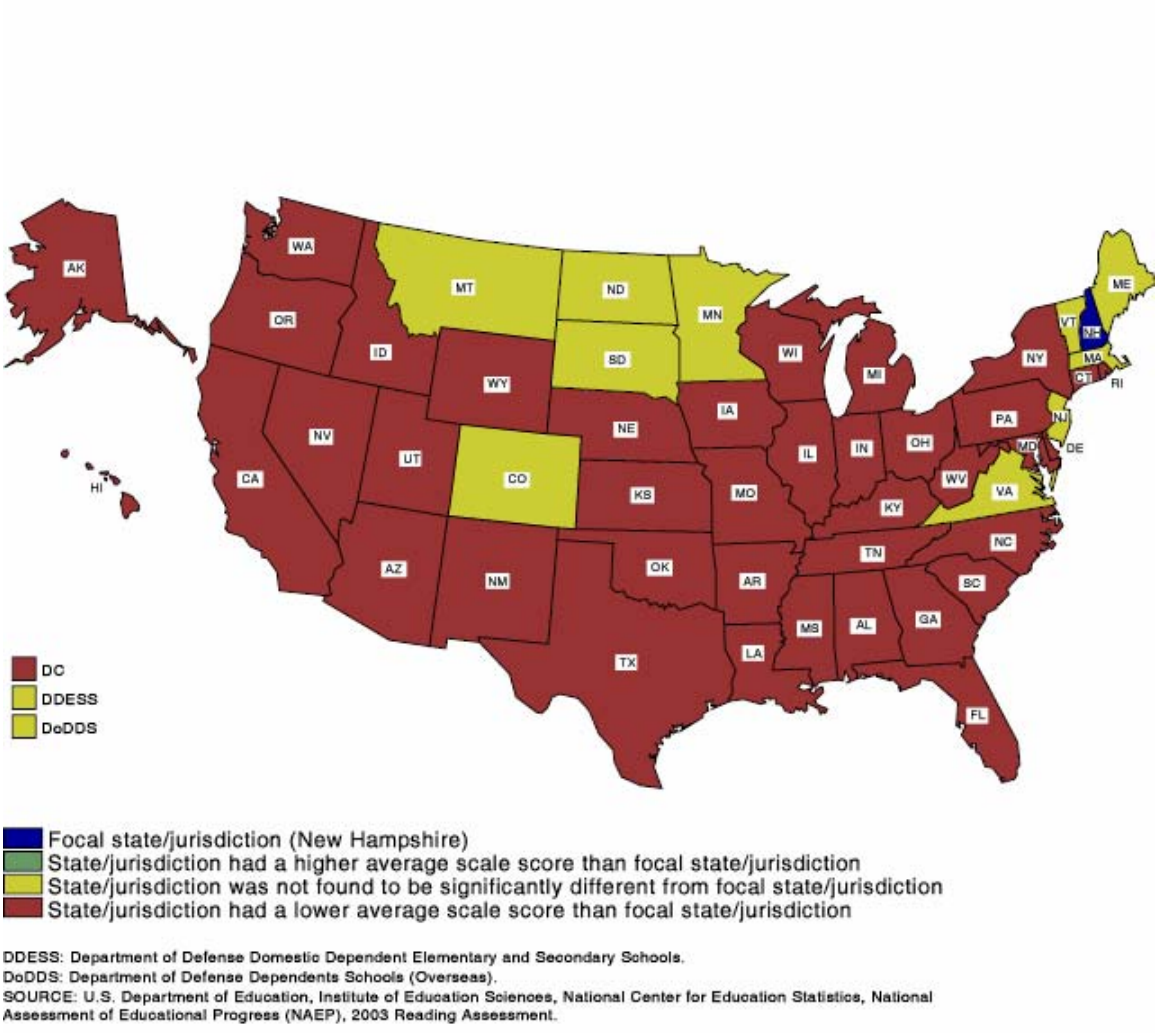


FIGURE 2

The Nation's Report Card 2003 State Assessment

New Hampshire's average reading scale score compared with scores for other participating jurisdictions, grade 8 public schools: 2003



DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools.
DoDDS: Department of Defense Dependents Schools (Overseas).
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.

Comparisons by Achievement Levels

Figure 3 permits comparisons of all jurisdictions participating in the NAEP 2003 reading assessment in terms of percentages of grade 8 students performing at or above the *Proficient* level. The participating states and jurisdictions are grouped into categories reflecting student performance compared to that in New Hampshire. The jurisdictions are grouped by whether the percentage of their students with scores at or above the *Proficient* level (including *Advanced*) was found to be higher than, not significantly different from, or lower than the percentage in New Hampshire. Note that the arrangement of the states and the other jurisdictions within each category is alphabetical; statistical comparisons among jurisdictions within each of the three categories are not included in this report. Cross-state comparisons are available at http://nces.ed.gov/nationsreportcard/states/.

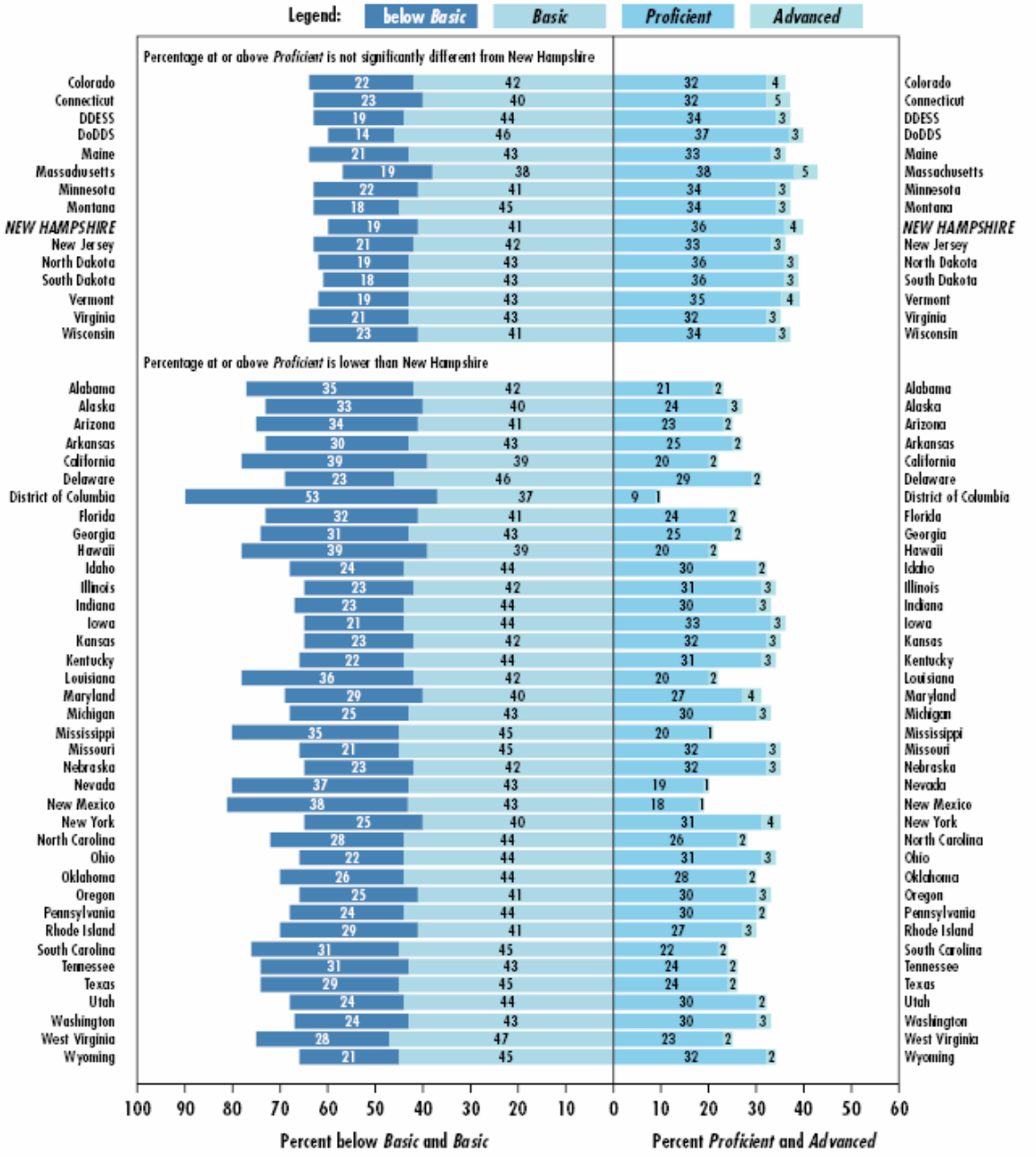
Grade 8 Achievement-Level Comparisons Results

- At grade 8, 14 jurisdictions had percentages of students at or above the *Proficient* level that were not significantly different from that of New Hampshire, and 38 jurisdictions had lower percentages than that of New Hampshire.

FIGURE 3

The Nation's Report Card 2003 State Assessment

Percentage of students within each reading achievement-level range, and New Hampshire's percentage at or above Proficient compared with other participating jurisdictions, grade 8 public schools: By state, 2003



DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools.
DoDDS: Department of Defense Dependents Schools (Overseas).
NOTE: The bars above contain percentages of students in each NAEP reading achievement range. Achievement levels corresponding to each population of students are aligned at the point where the *Proficient* category begins, so that they may be compared at *Proficient* and above. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.

Reading Performance by Demographic Characteristics

This section of the report presents results for students in New Hampshire and the nation by demographic characteristics. Student performance data are reported for:

- Gender
- Race/ethnicity
- Eligibility for free/reduced-price school lunch
- Type of location (2002 and later)

Definitions of NAEP reporting groups are available on the NAEP web site (http://nces.ed.gov/nationsreportcard/reading/results2003/interpret-results.asp#RepGroups).

Each of the variables is reported in tables that present the percentage of students belonging to each subgroup in the first column and the average scale score in the second column. The columns to the right show the percentage of students at or above each achievement-level.

The reader is cautioned against making causal inferences about the performance of groups of students relative to demographic variables. Many factors other than those discussed here, including home and school factors, may affect student performance.

NAEP collects information on many additional variables, including school and home factors related to achievement. All of this information is in an interactive database available on the NAEP web site (http://nces.ed.gov/nationsreportcard/naepdata/).

Gender

Information on student gender is reported by schools on rosters of students eligible to be assessed.

Table 3 shows scale scores and achievement-level data for public-school students at grade 8 in New Hampshire and the nation by gender. In 1998 only, results were obtained for student samples for which accommodations were permitted and were not permitted. However, in the text of this report, comparisons to 1998 results refer only to the sample in which accommodations were permitted.

Grade 8 Scale Score Results by Gender

- In New Hampshire, male students' average scale score was 265 in 2003. This was lower than that of female students (276).
- In 2003, male students in New Hampshire had an average scale score in Reading (265) that was higher than that of male students across the nation (256). Female students in New Hampshire had an average score (276) that was higher than that of female students nationwide (267).

Grade 8 Achievement-Level Results by Gender

- In 2003, 34 percent of males and 47 percent of females performed at or above the *Proficient* level in New Hampshire. The difference between these percentages was significant.
- The percentage of males in New Hampshire's public schools who were at or above the *Proficient* level in 2003 (34 percent) was greater than that of males in the nation (25 percent).
- The percentage of females in New Hampshire's public schools who were at or above the *Proficient* level in 2003 (47 percent) was greater than that of females in the nation (35 percent).

TABLE 3

The Nation's Report Card 2003 State Assessment

Average reading scale scores and percentage of students at or above each achievement level, by gender, grade 8 public schools: 2003

	Percentage of Students	Average Scale Score	Below *Basic*	At or above *Basic*	At or above *Proficient*	At *Advanced*
Male						
Accommodations permitted						
2003 New Hampshire	49 (0.9)	265 (1.3)	24 (1.6)	76 (1.6)	34 (1.9)	2 (0.6)
Nation (Public)	50 (0.2)	256 (0.3)	33 (0.3)	67 (0.3)	25 (0.3)	2 (0.1)
Female						
Accommodations permitted						
2003 New Hampshire	51 (0.9)	276 (1.2)	14 (1.2)	86 (1.2)	47 (2.2)	5 (0.6)
Nation (Public)	50 (0.2)	267 (0.3)	23 (0.3)	77 (0.3)	35 (0.3)	4 (0.1)

NOTE: The NAEP reading scale ranges from 0 to 500. The standard errors of the statistics in the table appear in parentheses. Achievement levels correspond to the following points on the NAEP reading scale: below *Basic*, 242 or lower; *Basic*, 243-280; *Proficient*, 281-322; and *Advanced*, 323 and above. All differences were tested for statistical significance at the 0.05 level using unrounded numbers. Details may not sum to totals due to rounding. Performance comparisons may be affected by differences in exclusion rates for students with disabilities and limited-English-proficient students in the NAEP samples and changes in sample sizes. NAEP sample sizes have increased since 2002 compared to previous years, resulting in smaller detectable differences than in previous assessments.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.

Race/Ethnicity

Schools report the racial/ethnic subgroup that best described the students eligible to be assessed. The five mutually exclusive categories are White, Black, Hispanic, Asian/Pacific Islander, and American Indian/Alaska Native.

Table 4 shows scale scores and achievement-level data for public-school students at grade 8 in New Hampshire and the nation by race/ethnicity. In 1998 only, results were obtained for student samples for which accommodations were permitted and were not permitted. However, in the text of this report, comparisons to 1998 results refer only to the sample in which accommodations were permitted.

Grade 8 Scale Score Results by Race/Ethnicity

The sample size was not sufficient to permit a reliable estimate for Black, Hispanic, Asian/Pacific Islands, and American/Alaska Native groups.

Grade 8 Achievement-Level Results by Race/Ethnicity

The sample size was not sufficient to permit a reliable estimate for Black, Hispanic, Asian/Pacific Islands, and American/Alaska Native groups.

TABLE 4

The Nation's Report Card 2003 State Assessment

Average reading scale scores and percentage of students at or above each achievement level, by race/ethnicity, grade 8 public schools: 2003

	Percentage of Students	Average Scale Score	Below *Basic*	At or above *Basic*	At or above *Proficient*	At *Advanced*
White						
Accommodations permitted						
2003 New Hampshire	94 (0.5)	272 (0.9)	18 (1.1)	82 (1.1)	41 (1.5)	4 (0.4)
Nation (Public)	61 (0.4)	270 (0.2)	18 (0.3)	82 (0.3)	39 (0.3)	4 (0.1)
Black						
Accommodations permitted						
2003 New Hampshire	2 (0.3)	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	17 (0.3)	244 (0.5)	47 (0.6)	53 (0.6)	12 (0.4)	# (0.1)
Hispanic						
Accommodations permitted						
2003 New Hampshire	2 (0.3)	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	15 (0.3)	244 (0.7)	46 (1.0)	54 (1.0)	14 (0.6)	1 (0.2)
Asian/Pacific Islander						
Accommodations permitted						
2003 New Hampshire	1 (0.2)	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	4 (0.2)	268 (1.2)	22 (1.3)	78 (1.3)	38 (1.7)	5 (0.6)
American Indian						
Accommodations permitted						
2003 New Hampshire	# (0.1)!	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	1 (0.1)	248 (1.7)	41 (2.4)	59 (2.4)	18 (1.6)	1 (0.3)

--- Reporting standards are not met. Sample size is insufficient to permit a reliable estimate.
Estimate rounds to zero.
! The nature of the sample does not allow accurate determination of the variability of the statistic.
NOTE: The NAEP reading scale ranges from 0 to 500. The standard errors of the statistics in the table appear in parentheses. Achievement levels correspond to the following points on the NAEP reading scale: below *Basic*, 242 or lower; *Basic*, 243-280; *Proficient*, 281-322; and *Advanced*, 323 and above. All differences were tested for statistical significance at the 0.05 level using unrounded numbers. Details may not sum to totals due to rounding. Performance comparisons may be affected by differences in exclusion rates for students with disabilities and limited-English-proficient students in the NAEP samples and changes in sample sizes. NAEP sample sizes have increased since 2002 compared to previous years, resulting in smaller detectable differences than in previous assessments.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.

Free/Reduced-Price Lunch Eligibility

NAEP collects data on eligibility for the federal program providing free or reduced-price school lunches. The free/reduced-price lunch component of the National School Lunch Program (NSLP) offered through the U.S. Department of Agriculture (USDA) is designed to ensure that children near or below the poverty line receive nourishing meals. This program is available to public schools, nonprofit private schools, and residential child-care institutions. Eligibility is determined through the USDA's Income Eligibility Guidelines, and results for this category of students are included as an indicator of poverty. NAEP first collected information on participation in this program in 1996.

Table 5 shows scale scores and achievement-level data for public-school students at grade 8 in New Hampshire and the nation by eligibility for free/reduced-price lunch. In 1998 only, results were obtained for student samples for which accommodations were permitted and were not permitted. However, in the text of this report, comparisons to 1998 results refer only to the sample in which accommodations were permitted.

Grade 8 Scale Score Results by Free/Reduced-Price Lunch Eligibility

- Students in New Hampshire eligible for free/reduced-price lunch had an average Reading scale score of 255. This was lower than that of students in New Hampshire not eligible for this program (273).

- Students in New Hampshire eligible for free/reduced-price lunch had an average scale score (255) that was higher than that of students in the nation who were eligible (246).

Grade 8 Achievement-Level Results by Free/Reduced-Price Lunch Eligibility

- In New Hampshire, 22 percent of students who were eligible for free/reduced-price lunch and 43 percent of those who were not eligible for this program performed at or above the *Proficient* level. These percentages were found to be significantly different from one another.

- For students in New Hampshire who were eligible for free/reduced-price lunch, the percentage at or above the *Proficient* level (22 percent) was greater than the corresponding percentage for their counterparts around the nation (15 percent).

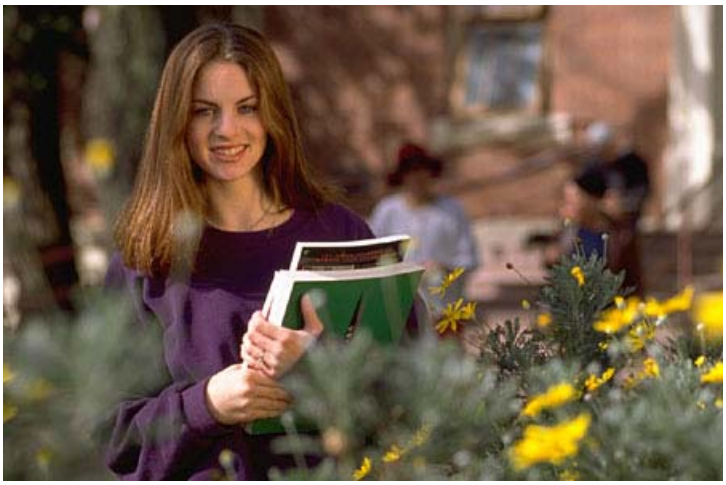
TABLE 5

The Nation's Report Card 2003 State Assessment

Average reading scale scores and percentage of students at or above each achievement level, by eligibility for free/reduced-price school lunch, grade 8 public schools: 2003

	Percentage of Students	Average Scale Score	Below *Basic*	At or above *Basic*	At or above *Proficient*	At *Advanced*
Eligible						
Accommodations permitted						
2003 New Hampshire	14 (1.0)	255 (2.9)	34 (3.7)	66 (3.7)	22 (2.4)	3 (0.9)
Nation (Public)	36 (0.4)	246 (0.4)	44 (0.5)	56 (0.5)	15 (0.3)	1 (0.1)
Not Eligible						
Accommodations permitted						
2003 New Hampshire	79 (1.1)	273 (0.9)	17 (1.1)	83 (1.1)	43 (1.8)	4 (0.5)
Nation (Public)	58 (0.5)	271 (0.3)	18 (0.3)	82 (0.3)	39 (0.4)	4 (0.1)
Information Not Available						
Accommodations permitted						
2003 New Hampshire	7 (0.8)	278 (2.9)	15 (3.0)	85 (3.0)	49 (4.7)	6 (2.1)
Nation (Public)	6 (0.4)	262 (1.0)	28 (1.1)	72 (1.1)	31 (1.1)	3 (0.5)

NOTE: The NAEP reading scale ranges from 0 to 500. The standard errors of the statistics in the table appear in parentheses. Achievement levels correspond to the following points on the NAEP reading scale: below *Basic*, 242 or lower; *Basic*, 243-280; *Proficient*, 281-322; and *Advanced*, 323 and above. All differences were tested for statistical significance at the 0.05 level using unrounded numbers. Details may not sum to totals due to rounding. Performance comparisons may be affected by differences in exclusion rates for students with disabilities and limited-English-proficient students in the NAEP samples and changes in sample sizes. NAEP sample sizes have increased since 2002 compared to previous years, resulting in smaller detectable differences than in previous assessments.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.



Type of Location

Schools that participated in the assessment were classified into three mutually exclusive types of community in which the school is located: central city, urban fringe/large town, and rural/small town. These categories indicate the geographic locations of schools. Central city is geographical term meaning the largest city of a Metropolitan Statistical Area and is not synonymous with "inner city."

Recently, the National Center for Education Statistics (NCES) introduced new methods to identify the type of location assigned to each school in the Common Core of Data (CCD). The new methods were put into place by NCES in order to improve the quality of the assignments, and they take into account more information about the exact physical location of the school. The variable was revised in NAEP beginning with the 2000 assessment; therefore, results are not presented for assessment years prior to 2000.

Table 6 shows scale scores and achievement-level data for public-school students at grade 8 in New Hampshire and the nation by type of location.

Grade 8 Scale Score Results by Type of Location

- In 2003, in New Hampshire, the average scale score of students attending schools in central cities was lower than those of students in urban fringes/large towns and rural areas/small towns.

Grade 8 Achievement-Level Results by Type of Location

- In 2003, the percentage of students attending schools in central cities in New Hampshire who performed at or above the *Proficient* level was smaller than the corresponding percentages for students in urban fringes/large towns and rural areas/small towns.

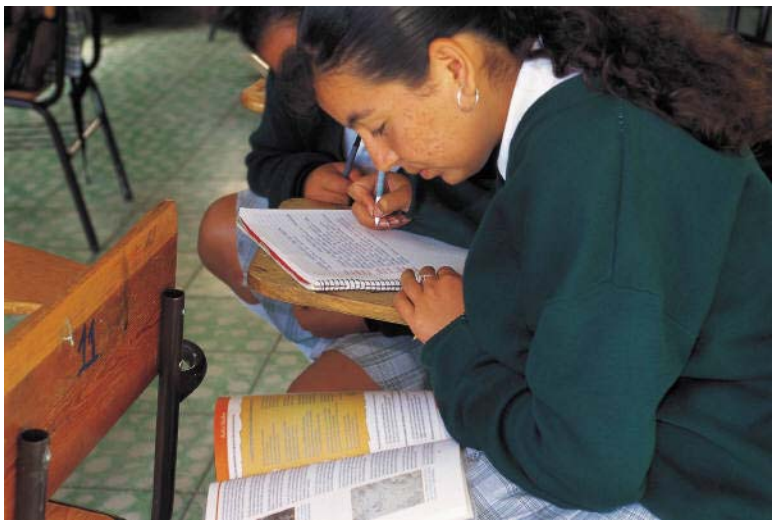


TABLE 6

The Nation's Report Card 2003 State Assessment

Average reading scale scores and percentage of students at or above each achievement level, by type of location, grade 8 public schools: 2003

	Percentage of Students	Average Scale Score	Below *Basic*	At or above *Basic*	At or above *Proficient*	At *Advanced*
Central City						
Accommodations permitted						
2003 New Hampshire	15 (0.8)	261 (3.1)	29 (3.4)	71 (3.4)	31 (3.2)	3 (1.1)
Nation (Public)	27 (0.4)	253 (0.5)	37 (0.6)	63 (0.6)	22 (0.5)	2 (0.1)
Urban Fringe/Large Town						
Accommodations permitted						
2003 New Hampshire	25 (0.6)	274 (1.2)	15 (1.7)	85 (1.7)	43 (2.6)	4 (0.8)
Nation (Public)	42 (0.4)	265 (0.5)	24 (0.4)	76 (0.4)	34 (0.5)	3 (0.1)
Rural/Small Town						
Accommodations permitted						
2003 New Hampshire	60 (1.2)	272 (1.3)	18 (1.4)	82 (1.4)	42 (2.0)	4 (0.6)
Nation (Public)	31 (0.4)	264 (0.4)	25 (0.4)	75 (0.4)	31 (0.5)	2 (0.1)

NOTE: The NAEP reading scale ranges from 0 to 500. The standard errors of the statistics in the table appear in parentheses. Achievement levels correspond to the following points on the NAEP reading scale: below *Basic*, 242 or lower; *Basic*, 243-280; *Proficient*, 281-322; and *Advanced*, 323 and above. All differences were tested for statistical significance at the 0.05 level using unrounded numbers. Details may not sum to totals due to rounding. Performance comparisons may be affected by differences in exclusion rates for students with disabilities and limited-English-proficient students in the NAEP samples and changes in sample sizes. NAEP sample sizes have increased since 2002 compared to previous years, resulting in smaller detectable differences than in previous assessments.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.

Toward a More Inclusive NAEP

NAEP endeavors to assess all students selected in the randomized sampling process, including students with disabilities (SD) as well as students who are classified by their schools as limited-English-proficient (LEP). Some students sampled for participation in NAEP can be excluded from the sample according to carefully defined criteria. School personnel, guided by the student's Individualized Education Program (IEP), as well as eligibility for Section 504 services, make decisions regarding inclusion in the assessment of students with disabilities. They also make decisions regarding inclusion of LEP students, based on NAEP's guidelines. This includes evaluating the student's capability of participating in the assessment in English, as well as taking into consideration the number of years the student has been receiving instruction in English.

Percentages of students excluded from NAEP may vary considerably across states, and within a state, across years. Comparisons of results across states and within a state across years should be interpreted with caution if the exclusion rates vary widely. The percentages of students classified as SD or LEP in all participating states and jurisdictions are available in an interactive database at the NAEP web site (http://nces.ed.gov/nationsreportcard/naepdata/).

The results displayed in this report and in other publications of the NAEP 2003 reading results are based on representative samples that include SD and LEP students who were assessed either with or without accommodations, based on NAEP's guidelines. Prior to 1998, however, in state NAEP reading assessments no testing accommodations or adaptations were made available to the special-needs students in the samples that served as the basis for reported results.

In the 1998 national and state reading assessments and the 2000 national (grade 4 only) reading assessment, NAEP drew a second representative sample of schools. Accommodations were made available for students in this sample who required them, provided the accommodation did not change the nature of what was tested. For example, students could be assessed one-on-one or in small groups, receive extended time, or use a large-print test book. However, for reading students were not permitted to have passages or test items read aloud. NAEP has used these comparable samples to study the effects of allowing accommodations for special-needs students in the assessments. A series of technical research papers covering various NAEP subject areas has been published with the results of these comparisons (see http://nces.ed.gov/nationsreportcard/about/inclusion.asp#research).

Table 7 displays the percentages of special-needs students identified, excluded, and accommodated at grade 8.

Table 8 presents the total number of students assessed, the percentage of students sampled that were excluded, and average scale scores for all participating states and other jurisdictions at grade 8.

TABLE 7

The Nation's Report Card 2003 State Assessment

Percentage of SD and LEP students in reading assessments identified, excluded, and assessed, grade 8 public schools: 2003

	SD and/or LEP		SD		LEP	
	New Hampshire	Nation (Public)	New Hampshire	Nation (Public)	New Hampshire	Nation (Public)
Accommodations permitted						
2003 Identified	19 (0.8)	19 (0.2)	18 (0.8)	14 (0.2)	2 (0.3)	6 (0.2)
Excluded	3 (0.3)	5 (0.1)	3 (0.3)	4 (0.1)	# (0.1)	2 (0.1)
Assessed under standard conditions	6 (0.5)	8 (0.2)	6 (0.5)	5 (0.1)	1 (0.2)	4 (0.2)
Assessed with accommodations	9 (0.8)	5 (0.1)	9 (0.8)	5 (0.1)	1 (0.2)	1 (0.1)

Estimate rounds to zero.
SD: Students with Disabilities. LEP: Limited-English-proficient students.
NOTE: The standard errors of the statistics in the table appear in parentheses. Detail may not sum to totals because of rounding. Some students were identified as both SD and LEP. Such students would be included in both the SD and LEP portions of the table.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.

TABLE 8

The Nation's Report Card 2003 State Assessment

Total number of students assessed, percentage of students sampled that were excluded, and average reading scale scores, grade 8 public schools: By state, 2003.

	Grade 8		
	Number assessed	Percentage excluded	Average scale score
Alabama	2,585	3 (0.4)	253 (1.5)
Alaska	2,498	2 (0.4)	256 (1.1)
Arizona	2,625	6 (0.8)	255 (1.4)
Arkansas	2,575	5 (0.5)	258 (1.3)
California	5,510	4 (0.5)	251 (1.3)
Colorado	2,710	3 (0.4)	268 (1.2)
Connecticut	2,725	4 (0.4)	267 (1.1)
Delaware	2,496	9 (0.5)	265 (0.7)
Florida	2,443	6 (0.7)	257 (1.3)
Georgia	4,219	3 (0.4)	258 (1.1)
Hawaii	2,768	5 (0.4)	251 (0.9)
Idaho	2,642	4 (0.4)	264 (0.9)
Illinois	4,039	5 (0.6)	266 (1.0)
Indiana	2,642	4 (0.5)	265 (1.0)
Iowa	2,823	5 (0.6)	268 (0.8)
Kansas	2,916	4 (0.4)	266 (1.5)
Kentucky	2,800	7 (0.6)	266 (1.3)
Louisiana	2,308	6 (0.6)	253 (1.6)
Maine	2,882	5 (0.4)	268 (1.0)
Maryland	2,449	3 (0.6)	262 (1.4)
Massachusetts	3,770	4 (0.6)	273 (1.0)
Michigan	2,625	6 (0.6)	264 (1.8)
Minnesota	2,605	3 (0.3)	268 (1.1)
Mississippi	2,694	5 (0.6)	255 (1.4)
Missouri	2,651	8 (0.8)	267 (1.0)
Montana	2,581	5 (0.4)	270 (1.0)
Nebraska	2,476	5 (0.4)	266 (0.9)
Nevada	2,651	4 (0.4)	252 (0.8)
New Hampshire	2,868	3 (0.3)	271 (0.9)
New Jersey	2,866	3 (0.6)	268 (1.2)
New Mexico	3,061	8 (1.3)	252 (0.9)
New York	3,424	7 (0.6)	265 (1.3)
North Carolina	4,057	7 (0.6)	262 (1.0)
North Dakota	2,612	4 (0.5)	270 (0.8)
Ohio	3,414	6 (0.7)	267 (1.3)
Oklahoma	2,839	4 (0.6)	262 (0.9)
Oregon	2,561	6 (0.8)	264 (1.2)
Pennsylvania	2,792	2 (0.4)	264 (1.2)
Rhode Island	2,643	4 (0.4)	261 (0.7)
South Carolina	2,446	8 (0.7)	258 (1.3)
South Dakota	2,770	3 (0.4)	270 (0.8)
Tennessee	2,655	3 (0.3)	258 (1.2)
Texas	4,378	8 (0.7)	259 (1.1)
Utah	2,732	3 (0.5)	264 (0.8)
Vermont	2,682	4 (0.4)	271 (0.8)
Virginia	2,733	9 (0.9)	268 (1.1)
Washington	2,625	4 (0.5)	264 (0.9)
West Virginia	2,234	9 (0.9)	260 (1.0)
Wisconsin	2,566	5 (0.6)	266 (1.3)
Wyoming	2,763	2 (0.3)	267 (0.5)
DC	1,922	8 (0.5)	239 (0.8)
DoDEA/DDESS	687	3 (0.6)	269 (1.4)
DoDEA/DoDDS	2,298	1 (0.2)	273 (0.7)

NOTE: The NAEP reading scale ranges from 0 to 500. The standard errors of the statistics in the table appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.

Appendix
Overview of Procedures Used for the NAEP 2003 Reading Assessment

The NAEP 2003 Reading Assessment

The National Assessment Governing Board (NAGB), created by Congress in 1988, is responsible for formulating policy for NAEP. NAGB is specifically charged with developing assessment objectives and test specifications. The design of the NAEP 2003 reading assessment follows the guidelines first provided in the framework developed for the 1992 assessment.[1] The framework underlying the 1992, 1994, 1998, 2000 (fourth grade only), 2002, and 2003 reading assessments reflects the expert opinions of educators and researchers about reading. The development of this framework and the specifications that guided the development of the assessment involved the critical input of hundreds of individuals across the country, including representatives of national education organizations, teachers, parents, policymakers, business leaders, and the interested general public. The framework development process was managed by the Council of Chief State School Officers (CCSSO) for NAGB.

The framework sets forth a broad definition of "reading literacy" that includes developing a general understanding of written text, thinking about it, and using various texts for many different purposes. In addition, the framework views reading as an interactive and constructive process involving the reader, the text, and the context of the reading experience. For example, readers may read stories to enjoy and appreciate the human experience, study science texts to form new hypotheses about knowledge, or follow directions to fill out a form. NAEP reflects current definitions of literacy by differentiating among three contexts for reading and four aspects of reading. The contexts for reading and aspects of reading make up the foundation of the NAEP reading assessment.

The "contexts for reading" dimension of the NAEP reading framework provides guidance for the types of texts to be included in the assessment. Although many commonalities exist among different types of reading contexts, different contexts do lead to real differences in what readers do. For example, when reading for literary experience, readers make complex, abstract summaries and identify major themes. They describe the interactions of various literary elements (e.g., setting, plot, characters, and theme). When reading for information, readers critically judge the form and content of the text and explain their judgments. They also look for specific pieces of information. When reading to perform a task, readers search quickly for specific pieces of information.

The "aspects of reading" dimension of the NAEP reading framework provides guidance for the types of comprehension questions to be included in the assessment. The four aspects are 1) forming a general understanding, 2) developing interpretation, 3) making reader/text connections, and 4) examining content and structure. These four aspects represent different ways in which readers develop understanding of a text. In forming a general understanding, readers must consider the text as a whole and provide a global understanding of it. As readers engage in developing interpretation, they must extend initial impressions in order to develop a more complete understanding of what was read. This involves linking information across parts of a text or focusing on specific information. When making reader/text connections, the reader must connect information in the text with knowledge and experience. This might include applying ideas in the text to the real world. Finally, examining content and structure requires critically evaluating, comparing and contrasting, and understanding the effect of different text features and authorial devices.

The following figure demonstrates the relationship between these reading contexts and aspects of reading in the NAEP reading assessment. Included in the figure are sample questions that illustrate how each aspect of reading is assessed within each reading context. (Note that reading to perform a task is not assessed at grade 4.)

Sample NAEP questions, by aspects of reading and contexts for reading specified in the reading framework

	Aspect of Reading			
Context for Reading	Forming a general understanding	Developing interpretation	Making reader/text connections	Examining content and structure
Reading for literary experience	What is the story/plot about?	How did this character change from the beginning to the end of the story?	What other character that you have read about had a similar problem?	What is the mood of this story and how does the author use language to achieve it?
Reading for information	What point is the author making about this topic?	What caused this change?	What other event in history or recent news is similar to this one?	Is this author biased? Support your answer with information about this article.
Reading to perform a task	What time can you get a nonstop flight to X?	What must you do before step 3?	Describe a situation in which you would omit step 5.	Is the information in this brochure easy to use?

SOURCE: National Assessment Governing Board. (2002). *Reading Framework for the 2003 National Assessment of Educational Progress*. Washington, DC: Author.

The assessment framework specifies not only the particular dimensions of reading literacy to be measured, but also the percentage of assessment questions that should be devoted to each. The target percentage distribution for contexts of reading and aspects of reading as specified in the framework, along with the actual percentage distribution in the assessment, are presented in the following tables.

Target and actual percentage distribution of questions, by context for reading, grades 4 and 8: 2003

		Context for Reading		
		Reading for literary experience	**Reading for information**	**Reading to perform a task**
Grade 4				
	Target	55	45	†
	Actual	50	50	†
Grade 8				
	Target	40	40	20
	Actual	28	41	30

† Not applicable. Reading to perform a task was not assessed at grade 4.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.

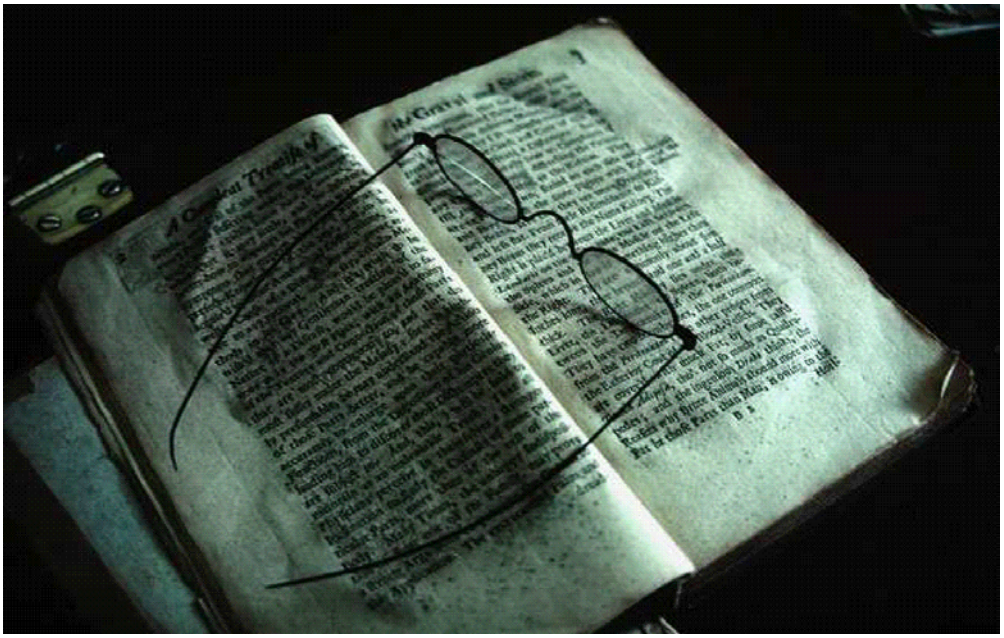
Target and actual percentage distribution of student time, by aspect of reading, grades 4 and 8: 2003

		Aspect of Reading		
		Forming a general understanding/ developing interpretation	**Making reader/text connections**	**Examining content and structure**
Grade 4				
	Target	60	15	25
	Actual	61	17	22
Grade 8				
	Target	55	15	30
	Actual	56	18	26

NOTE: Actual percentages are based on the classifications agreed upon by NAEP's Instrument Development Panel. It is recognized that making discrete classifications for these categories is difficult and that independent efforts to classify questions have led to different results.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment.

The actual content of the assessment has varied from the targeted distribution. For example, at grade 8 reading for literary experience falls below the target proportions and reading for information falls above the target proportions specified in the framework. The reading instrument development panel overseeing the development of the assessment recognized this variance but felt strongly that assessment questions must be sensitive to the unique elements of the authentic reading materials being used. Thus, the distribution of question classifications will vary across reading passages and reading contexts.

1. National Assessment Governing Board. (2002). *Reading Framework for the 2003 National Assessment of Educational Progress*. Washington, DC: Author.



The Assessment Design

Each student who participated in the NAEP 2003 reading assessment received a booklet containing three or four sections: a set of general background questions, a set of subject-specific background questions, and one or two sets of questions assessing students' comprehension of a text or texts. The sets of questions assessing students' comprehension are referred to as "blocks." Each block contains one or more reading passages and a set of comprehension questions. At grade 8, students were given either two 25-minute blocks or one 50-minute block. At grade 4, however, only 25-minute blocks were used.

The blocks contain a combination of multiple-choice and constructed-response questions. Multiple-choice questions require students to select the best answer from a set of four options. Constructed-response questions require students to provide their own written response to an open-ended question. Short constructed-response questions may require a response of only a sentence or two for the answer to be considered complete. Extended constructed-response questions, however, may require a response of a paragraph or more for the answer to receive full credit. Each constructed-response question has its own unique scoring guide that is used by trained scorers to rate students' responses.

The grade 4 assessment consisted of ten 25-minute blocks: five blocks of literary texts and questions and five blocks of informative texts and questions. Each block contained at least one passage corresponding to one of the contexts for reading and 9–12 multiple-choice and constructed-response questions. In most blocks, one of the constructed-response questions required an extended response. As a whole, the 2003 fourth-grade assessment consisted of 49 multiple-choice questions, 45 short constructed-response questions, and 8 extended constructed-response questions.

The grade 8 assessment consisted of twelve 25-minute blocks (four literary, four informative, and four to perform a task) and one 50-minute block (informative). Each block contained at least one passage corresponding to one of the contexts for reading and 9–13 multiple-choice and constructed-response questions. Most blocks contained at least one extended constructed-response question. As a whole, the eighth-grade assessment consisted of 58 multiple-choice questions, 68 short constructed-response questions, and 15 extended constructed-response questions.

The assessment design allowed maximum coverage of a range of reading abilities at each grade, while minimizing the time burden for any one student. This was accomplished through the use of matrix sampling of items in which representative samples of students took various portions of the entire pool of assessment questions. Individual students are required to take only a small portion, but the aggregate results across the entire assessment allow broad reporting of reading abilities for the targeted population.

In addition to matrix sampling, the assessment design utilized a procedure for distributing blocks across booklets that controlled for position and context effects. Students receive different blocks of passages and comprehension questions in their booklets according to a procedure called "partially balanced incomplete block (pBIB) spiraling." This procedure assigned blocks of questions in a manner that balanced the positioning of blocks across booklets and balanced the pairing of blocks within booklets according to the context for reading. Blocks were balanced within each context for reading and were partially balanced across contexts for reading. The spiraling aspect of this procedure cycles the booklets for administration so that, typically, only a few students in any assessment session receive the same booklet.

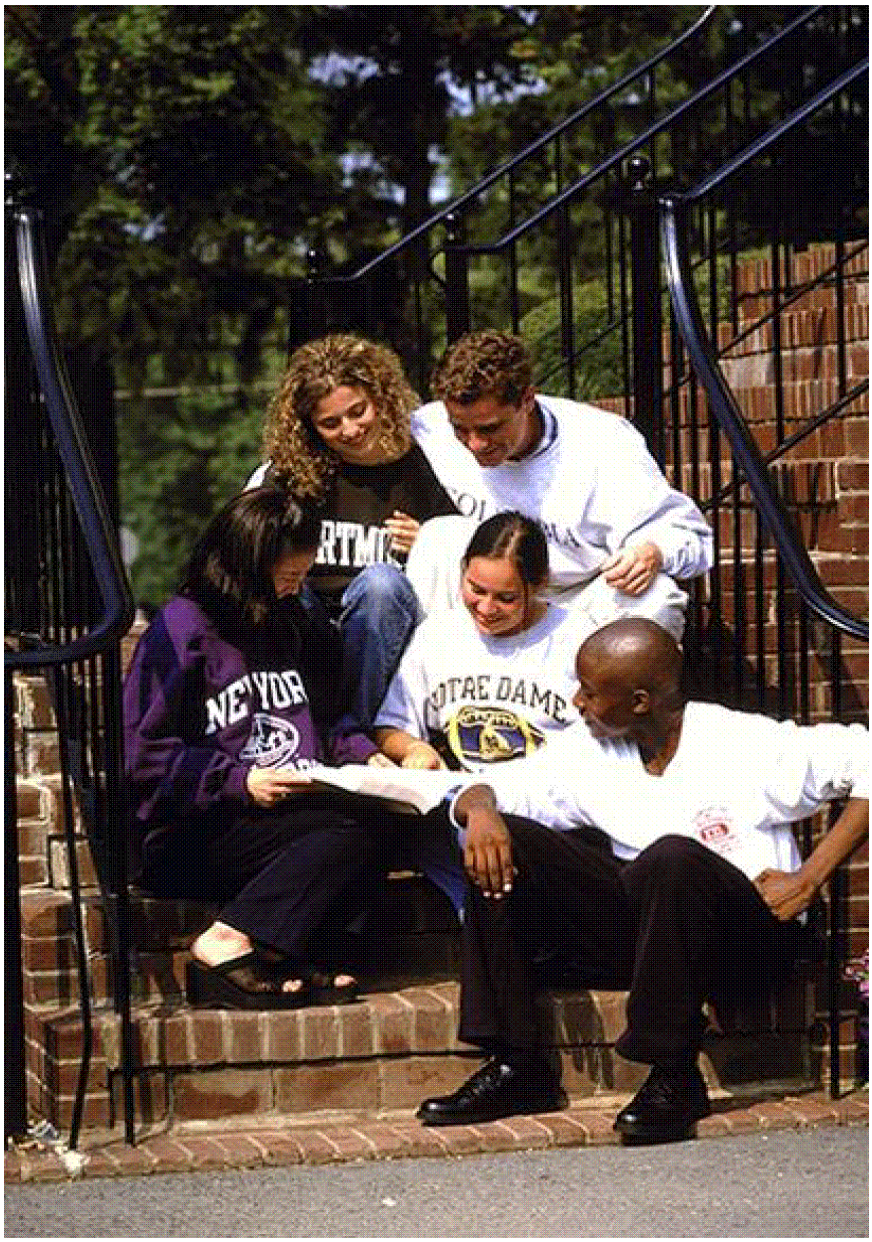
In addition to the student assessment booklets, three other instruments provided data relating to the assessment: a teacher questionnaire, a school questionnaire, and a questionnaire for students with disabilities and limited-English-proficient students (SD/LEP). The teacher questionnaire was administered to teachers of fourth- and eighth-grade students participating in the assessment and included questions about the teacher's background and classroom organization. The fourth-grade teacher questionnaire also included questions on reading instruction. The school questionnaire was given to the principal or other administrator in each participating school and included questions related to school characteristics, policies, programs, and the composition and background of the student body.

The SD/LEP questionnaire was completed by a school staff member knowledgeable about those students selected to participate in the assessment who were identified as having an Individualized Education Program (IEP) or equivalent plan (for reasons other than being gifted or talented) or having limited English proficiency. An SD/LEP questionnaire was completed for each identified student regardless of whether the student participated in the assessment. Each SD/LEP questionnaire took approximately three minutes to complete and asked about the student and the special-education programs in which he or she participated.

Data Collection and Scoring

The NAEP 2003 reading assessment was conducted from January to March 2003 by contractors to the U.S. Department of Education. Trained field staff from Westat conducted the data collection. Materials from the 2003 assessment were shipped to Pearson, where trained staff evaluated the responses to the constructed-response questions using scoring rubrics, or guides, prepared by Educational Testing Service (ETS). Each constructed-response question had a unique scoring guide that defined the criteria used to evaluate students' responses. Short constructed-response questions were scored as either acceptable or unacceptable, or were rated according to three-level guides that permitted partial credit. Extended constructed-response questions were evaluated with four-level guides.

For the 2003 reading assessment, 3,913,147 constructed responses were scored. This number includes rescoring to monitor interrater reliability. The within-year average percentage of exact agreement for the 2003 national reliability sample was 90 percent at both fourth and eighth grades.



Data Analysis and IRT Scaling

After the professional scoring, all information was transcribed into the NAEP database at ETS. Each processing activity was conducted with rigorous quality control. After the assessment information was compiled in the database, the data were weighted according to the population structure. The weighting for the national and state samples reflected the probability of selection for each student as a result of the sampling design, adjusted for nonresponse.[1]

Analyses were then conducted to determine the percentages of students who gave various responses to each cognitive and background question. In determining these percentages for the cognitive questions, a distinction was made between missing responses at the end of a block (i.e., missing responses after the last question the student answered) and missing responses before the last observed response. Missing responses before the last observed response were considered intentional omissions. In analysis, omitted responses to multiple-choice items were scored as fractionally correct.[2] Omitted responses for constructed-response items were placed into the lowest score category. Missing responses after the last observed response were considered "not reached" and treated as if the questions had not been presented to the student. In calculating response percentages for each question, only students classified as having been presented the question were included in the denominator of the statistic.

It is standard NAEP practice to treat all nonrespondents to the last question in a block as if they had not reached the question. For multiple-choice and short constructed-response questions, this practice produces a reasonable pattern of results in that the proportion reaching the last question is not dramatically smaller than the proportion reaching the next-to-last question. However, for reading blocks that ended with extended constructed-response questions, there may be extremely large drops in the proportion of students attempting some of the final questions. Therefore, for blocks ending with an extended constructed-response question, students who answered the next-to-last question but did not respond to the extended constructed-response question were classified as having intentionally omitted the last question.

Item Response Theory (IRT) was used to estimate average reading scale scores for the nation and for various subgroups of interest within the nation. IRT models the probability of answering a question in a certain way as a mathematical function of proficiency or skill. The main purpose of IRT analysis is to provide a common scale on which performance can be compared among groups such as those defined by characteristics, including gender and race/ethnicity, even when students receive different blocks of items. One desirable feature of IRT is that it locates items and students on this common scale. In contrast to classical test theory, IRT does not rely solely on the total number of correct item responses, but uses the particular patterns of student responses to items in determining the student location on the scale. As a result, adding items that function at a particular point on the scale to the assessment does not change the location of the students on the scale, even though students may respond correctly to more items. It does increase the relative precision with which students are measured, particularly those students whose scale locations are close to the additional items.

The results for 1992, 1994, 1998, 2000, 2002, and 2003 are presented on the NAEP composite reading scale, developed in 1992. For the NAEP 1992 reading assessment, a scale ranging from 0 to 500 was created to report performance for each reading context: literary and informative at grade 4; and literary, informative, and task oriented at grade 8. The scales summarize student performance across all three types of questions in the assessment (multiple-choice, short constructed-response, and extended constructed-response).

Each reading scale was initially based on the distribution of student performance across all three grades in the 1992 national assessment (grades 4, 8, and 12) and had an average of 250 and a standard deviation of 50. The composite scale was created as an overall measure of students' reading performance. This composite scale is a weighted average of the three separate scales for the reading contexts (two at grade 4). The weight for each reading context is proportional to the relative importance assigned to the reading context by the specifications developed through the consensus planning process and given in the framework.

In producing the reading scales, three distinct IRT models were used. Multiple-choice questions were scaled using the three-parameter logistic (3PL) model; short constructed-response questions rated as acceptable or unacceptable were scaled using the two-parameter logistic (2PL) model; and short constructed-response questions rated according to a three-level guide, as well as extended constructed-response questions rated on a four-level guide, were scaled using a generalized partial credit (GPC) model.[3] Developed by ETS and first used in 1992, the GPC model permits the scaling of questions scored according to multipoint rating schemes. The model takes full advantage of the information available from each of the student response categories used for these more complex constructed-response questions.

The reading scale is composed of three types of questions: multiple-choice, short constructed-response (scored either dichotomously or allowing for partial credit), and extended constructed-response (scored according to a partial-credit model). Unfortunately, the question of how much information different types of questions contribute to the reading scale has no simple answer. The information provided by a given question is determined by the IRT model used to scale the question. It is a function of the item parameters and varies by level of reading proficiency.[4] Thus, the answer to the query "How much information do the different types of questions provide?" will differ for each level of reading performance. When considering the composite reading scale, the answer is even more complicated. The reading data are scaled separately by the two contexts for reading (reading for literary experience and reading for information) for grade 4, and the three contexts for reading (reading for literary experience, reading for information, and reading to perform a task) for grade 8, resulting in two or three separate subscales at each grade. The composite scale is a weighted combination of these subscales. IRT information functions are only strictly comparable when the item parameters are estimated together. Because the composite scale is based on three separate estimation runs, there is no direct way to compare the information provided by the questions on the composite scale.

Because of the NAEP pBIB spiraling design, students do not receive enough questions about a specific topic to provide reliable information about individual performance. Traditional test scores for individual students, even those based on IRT, would result in misleading estimates of population characteristics, such as subgroup means and percentages of students at or above a certain scale-score level. However, it is NAEP's goal to estimate these population characteristics. NAEP's objectives can be achieved with methodologies that produce estimates of the population-level parameters directly, without the intermediary computation of estimates of individuals. This is accomplished using marginal estimation scaling model techniques for latent variables.[5] Under the assumptions of the scaling models, these population estimates will be consistent in the sense that the estimates approach the model-based population values as the sample size increases. This would not be the case for population estimates obtained by aggregating optimal estimates of individual performance.[6]

---

1. Weighting procedures are described more fully under the topic "Weighting and Variance Estimation."

2. Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems,* p. 229. Hillsdale, NJ: Lawrence Erlbaum Associates.

3. Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement, 16*(2), 159–176.

4. Donoghue, J. R. (1994). An Empirical Examination of the IRT Information of Polytomously Scored Reading Items Under the Generalized Partial Credit Model. *Journal of Educational Measurement, 31*(4), 295–311.

5. Mislevy, R. J., and Sheehan, K. M. (1987). Marginal Estimation Procedures. In A. E. Beaton (Ed.). *Implementing the New Design: The NAEP 1983–1984 Technical Report* (Report No. 15-TR-20), pp. 260–293. Princeton, NJ: Educational Testing Service.

6. For theoretical and empirical justification of the procedures employed, see Mislevy, R. J. (1988). Randomization Based Inferences About Latent Variables from Complex Samples. *Psychometrika, 56*(2), 177–196. For additional discussion, see Thomas, N. (1993). Asymptotic Corrections for Multivariate Posterior Moments with Factored Likelihood Functions. *Journal of Computational and Graphical Statistics*, *25*, 351-372. Also see Mazzeo, J., Donoghue, J. R., and Johnson, M. (under review). Marginal Estimation in NAEP: Current Operational Procedures and *AM*.

Weighting and Variance Estimation

A complex sampling design was used to select the students who were assessed. The properties of a sample selected through such a design could be very different from those of a simple random sample, in which every student in the target population has an equal chance of selection and in which the observations from different sampled students can be considered to be statistically independent of one another. Therefore, the properties of the sample for the data collection design were taken into account during the analysis of the assessment data.

One way that the properties of the sample design were addressed was by using sampling weights to account for the fact that the probabilities of selection were not identical for all students. All population and subpopulation characteristics based on the assessment data were estimated using sampling weights. These weights included adjustments for school and student nonresponse.

Prior to 2002, the national samples used weights that had been poststratified to the Census or Current Population Survey (CPS) totals for the populations being assessed. Due to concerns about the availability of appropriate targets for poststratification as a result of changes in the reporting of race in the 2000 census, nonpoststratified weights have been used in the analysis of national samples since 2002. The state NAEP samples have always been analyzed using non-poststratified weights, since there were no targets available from CPS to use in poststratification.

Not only must appropriate estimates of population characteristics be derived, but appropriate measures of the degree of uncertainty must be obtained for those statistics. Two components of uncertainty are accounted for in the variability of statistics based on student ability: 1) the uncertainty due to sampling only a relatively small number of students, and 2) the uncertainty due to sampling only a portion of the cognitive domain of interest. The first component accounts for the variability associated with the estimated percentages of students who had certain background characteristics or who answered a certain cognitive question correctly.

Because NAEP uses complex sampling procedures, conventional formulas for estimating sampling variability that assume simple random sampling are inappropriate. NAEP uses a jackknife replication procedure to estimate standard errors. The jackknife standard error provides a reasonable measure of uncertainty for any student information that can be observed without error. However, because each student typically responds to only a few questions within any theme of reading, the scale score for any single student would be imprecise. In this case, NAEP's marginal estimation methodology can be used to describe the performance of groups and subgroups of students. The estimate of the variance of the students' posterior scale score distributions (which reflect the imprecision due to lack of measurement accuracy) is computed. This component of variability is then included in the standard errors of NAEP scale scores.[1]

Typically, when the standard error is based on a small number of students or when the group of students is enrolled in a small number of schools, the amount of uncertainty associated with the estimation of standard errors may be quite large. Estimates of standard errors subject to a large degree of uncertainty are followed on the tables by the "!" symbol to indicate that the nature of the sample does not allow accurate determination of the variability of the statistic. In such cases, the standard errors—and any confidence intervals or significance tests involving these standard errors—should be interpreted cautiously.

The reader is reminded that, as with findings from all surveys, NAEP results are subject to other kinds of error, including the effects of imperfect adjustment for student and school nonresponse and unknowable effects associated with the particular instrumentation and data collection methods. Nonsampling errors can be attributed to a number of sources—inability to obtain complete information about all selected schools in the sample (some students or schools refused to participate, or students participated but answered only certain questions); ambiguous definitions; differences in interpreting questions; inability or unwillingness to give correct background information; mistakes in recording, coding, or scoring data; and other errors in collecting, processing, sampling, and estimating missing data. The extent of nonsampling errors is difficult to estimate and, because of their nature, the impact of such errors cannot be reflected in the data-based estimates of uncertainty provided in NAEP reports.

---

1. For further details, see Johnson, E. G., and Rust, K. F. (1992). Population Inferences and Variance Estimation for NAEP Data. *Journal of Educational Statistics, 17*(2), 175–190.

Drawing Inferences from the Results

The reported statistics are estimates and are therefore subject to a measure of uncertainty. There are two sources of such uncertainty. First, NAEP uses a sample of students rather than testing all students. Second, all assessments have some amount of uncertainty related to the fact that they cannot ask all questions that might be asked in a content area. The magnitude of this uncertainty is reflected in the standard error of each of the estimates. When the percentages or average scale scores of certain groups are compared, the estimated standard error should be taken into account. Therefore, the comparisons are based on statistical tests that consider the estimated standard errors of those statistics and the magnitude of the difference among the averages or percentages.

For the data from this report, all the estimates have corresponding estimated standard errors of the estimates. For example, the following tables show the average national public-school scale score for the NAEP 1992–2003 national assessments and achievement-level results. In both tables, estimated standard errors appear in parentheses next to each estimated scale score or percentage. For the estimated standard errors corresponding to other data from this report, the reader can go to the data tool on the NCES web site at http://nces.ed.gov/nationsreportcard/naepdata/.

Average reading scale scores and standard errors, grades 4 and 8 public schools: 1992–2003

	Accommodations not permitted				Accommodations permitted			
	1992	1994	1998	2000	1998	2000	2002	2003
Grade 4	215(1.0)	212(1.1)*	215(0.8)	215(0.9)*	215(1.0)	213(1.2)*	217(0.5)	216(0.3)
Grade 8	258(1.0)*	257(0.8)*	261(0.8)	---	261(0.8)	---	263(0.5)*	261(0.3)

--- Not available. Data were not collected at grade 8 in 2000
* Significantly different from 2003.
NOTE: Standard errors of the estimated scale scores appear in parentheses.
In addition to allowing for accommodations, the accommodations-permitted results at grade 4 (1998–2000) differ slightly from previous years, and from previous reported results for 1998 and 2000 due to changes in sample weighting procedures.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 1998, 2000, 2002, and 2003 Reading Assessments.

Percentage of students and standard errors, by reading achievement level, grades 4 and 8 public schools: 1992–2003

		Below *Basic*	At *Basic*	At or above *Basic*	At or above *Proficient*
Grade 4					
Accommodations not permitted	1992	40(1.1)	6(0.6)	60(1.1)	27(1.3)*
	1994	41(1.1)*	7(0.7)	59(1.1)*	28(1.2)
	1998	39(1.0)	6(0.5)	61(1.0)	29(0.9)
	2000	40(0.9)	7(0.6)	60(0.9)	30(1.0)
Accommodations permitted	1998	42(1.3)*	6(0.5)	58(1.3)*	28(1.0)*
	2000	43(1.5)*	6(0.6)	57(1.5)*	28(1.2)
	2002	38(0.5)	6(0.2)*	62(0.5)	30(0.5)
	2003	38(0.3)	7(0.1)	62(0.3)	30(0.3)
Grade 8					
Accommodations not permitted	1992	33(1.1)*	2(0.3)	67(1.1)*	27(1.1)*
	1994	33(0.9)*	2(0.3)	67(0.9)*	27(0.9)*
	1998	28(0.9)	2(0.4)	72(0.9)	31(0.9)
Accommodations permitted	1998	29(0.8)	2(0.3)	71(0.8)	30(1.1)
	2002	26(0.5)*	2(0.2)	74(0.5)*	31(0.6)
	2003	28(0.3)	3(0.1)	72(0.3)	30(0.1)

* Significantly different from 2003.
NOTE: Standard errors of the estimated percentages appear in parentheses.
Detail may not sum to totals because of rounding.
In addition to allowing for accommodations, the accommodations-permitted results at grade 4 (1998–2000) differ slightly from previous years, and from previous reported results for 1998 and 2000 due to changes in sample weighting procedures.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 1998, 2000, 2002, and 2003 Reading Assessments.

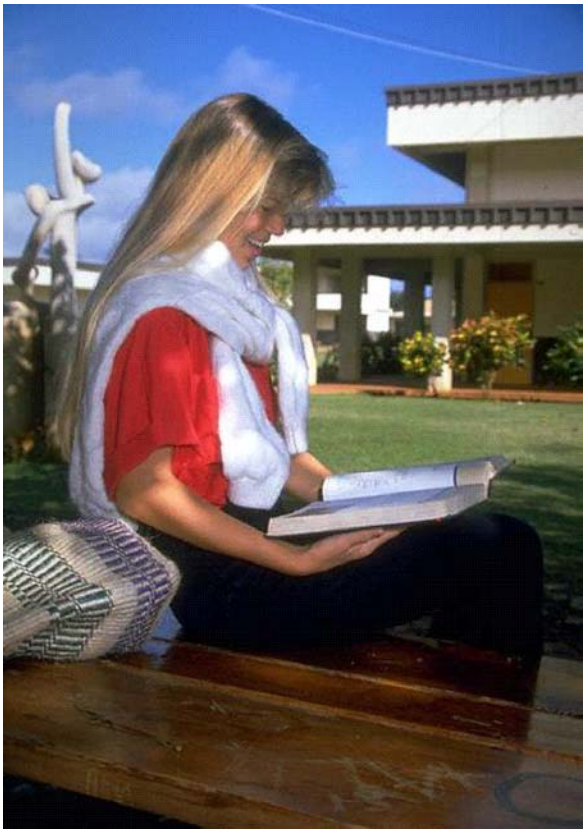
Using confidence intervals based on the standard errors provides a way to take into account the uncertainty associated with sample estimates and to make inferences about the population averages and percentages in a manner that reflects that uncertainty. An estimated sample average scale score plus or minus 1.96 standard errors approximates a 95 percent confidence interval for the corresponding population quantity. This statement means that one can conclude with an approximately 95 percent level of confidence that the average performance of the entire population of interest (e.g., all fourth-grade students in public schools) is within plus or minus 1.96 standard errors of the sample average.

For example, suppose that the average reading scale score of the students in a particular group was 256 with an estimated standard error of 1.2. An approximately 95 percent confidence interval for the population quantity would be as follows:

Average ± 1.96 standard errors
256 ± 1.96 x 1.2
256 ± 2.4
(253.6, 258.4)

Thus, one can conclude with a 95 percent level of confidence that the average scale score for the entire population of students in that group is between 253.6 and 258.4. It should be noted that this example and the examples in the following sections are illustrative. More precise estimates carried out to one or more decimal places are used in the actual analyses.

Similar confidence intervals can be constructed for percentages, if the percentages are not extremely large or extremely small. Extreme percentages should be interpreted with caution. Adding or subtracting the standard errors associated with extreme percentages could cause the confidence interval to exceed 100 percent or fall below 0 percent, resulting in numbers that are not meaningful.



Analyzing Group Differences in Averages and Percentages

Statistical tests determine whether, based on the data from the groups in the sample, there is strong enough evidence to conclude that the averages or percentages are actually different for those groups in the population. If the evidence is strong (i.e., the difference is statistically significant), the report describes the group averages or percentages as being different (e.g., one group performed higher or lower than another group), regardless of whether the sample averages or percentages appear to be approximately the same. The reader is cautioned to rely on the results of the statistical tests rather than on the apparent magnitude of the difference between sample averages or percentages when determining whether the sample differences are likely to represent actual differences among the groups in the population.

To determine whether a real difference exists between the average scale scores (or percentages of a certain attribute) for two groups in the population, one needs to obtain an estimate of the degree of uncertainty associated with the difference between the averages (or percentages) of these groups for the sample. This estimate of the degree of uncertainty, called the "standard error of the difference" between the groups, is obtained by taking the square of each group's standard error, summing the squared standard errors, and taking the square root of that sum.

$$\text{Standard Error of the Difference} = SE_{A-B} = \sqrt{(SE_A^2 + SE_B^2)}$$

The standard error of the difference can be used, just like the standard error for an individual group average or percentage, to help determine whether differences among groups in the population are real. The difference between the averages or percentages of the two groups plus or minus 1.96 standard errors of the difference represents an approximately 95 percent confidence interval. If the resulting interval includes zero, there is insufficient evidence to claim a real difference between the groups in the population. If the interval does not contain zero, the difference between the groups is statistically significant at the 0.05 level.

The following example of comparing groups addresses the problem of determining whether the average reading scale score of group A is higher than that of group B. The sample estimates of the average scale scores and estimated standard errors are as follows:

Group	Average Scale Score	Standard Error
A	218	0.9
B	216	1.1

The difference between the estimates of the average scale scores of groups A and B is two points (218–216). The estimated standard error of this difference is $\sqrt{(0.9^2 + 1.1^2)} = 1.4$

Thus, an approximately 95 percent confidence interval for this difference is plus or minus 1.96 standard errors of the difference:

$$2 \pm 1.96 \times 1.4$$
$$2 \pm 2.7$$
$$(-0.7, 4.7)$$

The value zero is within the confidence interval; therefore, there is insufficient evidence to conclude that group A outperformed group B.

The procedure above is appropriate to use when it is reasonable to assume that the groups being compared have been independently sampled for the assessment. Such an assumption is clearly warranted when comparing results across assessment years (e.g., comparing the 2002 and 2003 results for a particular state or subgroup) or when comparing results for one state with another. This is the approach used for NAEP reports when comparisons involving independent groups are made. The assumption of independence is violated to some degree when comparing group results for the nation or a particular state (e.g., comparing national 2003 results for males and females), since these samples of students have been drawn from the same schools. The impact of this violation of the independence assumption on the outcome of the statistical tests is assumed to be small when the groups being compared do not share students (as is the case, for example, comparing males and females), and NAEP, by convention, has, for computational convenience, routinely applied the procedures described above to those cases as well.

When making comparisons of results for groups that share a considerable proportion of students in common, it is not appropriate to ignore such dependencies. In such cases, NAEP has used procedures appropriate to comparing dependent groups. When the dependence in group results is due to the overlap in samples (e.g., when a subgroup is being compared to a total group), a simple modification of the usual standard error of the difference formula can be used. The formula for such cases is $SE_{Total-Subgroup} = \sqrt{(SE^2_{Total} + SE^2_{Subgroup} - 2pSE^2_{Subgroup})}$ where *p* is the proportion of the total group contained in the subgroup.[1] This formula was used for this report when a state was compared to the aggregate nation or a school district was compared to the entire state it belongs to.

---

1. This is a special form of the common formula for standard error of dependent samples. The standard formula can be found, for example, in Kish, L. (1995). *Survey Sampling.* New York: John Wiley and Sons, Inc.

Conducting Multiple Tests

The procedures used to determine whether group differences in the samples represent actual differences among the groups in the population and the certainty ascribed to intervals (e.g., a 95 percent confidence interval) are based on statistical theory that assumes that only one confidence interval or test of statistical significance is being performed. However, there are times when many different groups are being compared (i.e., multiple sets of confidence intervals are being analyzed). In sets of confidence intervals, statistical theory indicates that the certainty associated with the entire set of intervals is less than that attributable to each individual comparison from the set. To hold the significance level for the set of comparisons at a particular level (e.g., 0.05), standard methods must be adjusted by multiple comparison procedures.[1] One such procedure, the Benjamini-Hochberg False Discovery Rate (FDR) procedure, was used to control the certainty level.[2]

Unlike other multiple comparison procedures that control the familywise error rate (i.e., the probability of making even one false rejection in the set of comparisons), the FDR procedure controls the expected proportion of falsely rejected hypotheses. Furthermore, the FDR procedure used in NAEP is considered appropriately less conservative than familywise procedures for large families of comparisons.[3] Therefore, the FDR procedure is more suitable for multiple comparisons in NAEP than other procedures.

To illustrate how the FDR procedure is used, consider the comparisons of current and previous years' average scale scores for the five groups presented in the following table. The test statistic shown is the difference in average scale scores divided by the estimated standard error of the difference. (Rounding of the data occurs after the test is done.)

Example of False Discovery Rate comparisons of average scale scores for different groups of students

	Previous year		Current year		Previous year and current year			
	Average scale score	Standard error	Average scale score	Standard error	Difference in averages	Standard error of difference	Test Statistic	Percent confidence*
Group 1	224	1.3	226	1.0	2.08	1.62	1.29	20
Group 2	187	1.7	193	1.7	6.31	2.36	2.68	1
Group 3	191	2.6	197	1.7	6.63	3.08	2.15	4
Group 4	229	4.4	232	4.6	3.24	6.35	0.51	62
Group 5	201	3.4	196	4.7	-5.51	5.81	-0.95	35

* The percent confidence is 2(1-F(x)), where F(x) is the cumulative distribution of the t-distribution with the degrees of freedom adjusted to reflect the complexities of the sample design.

The difference in average scale scores and its estimated standard error can be used to find an approximately 95 percent confidence interval or they can be used to identify a confidence percentage. The confidence percentage for the test statistics is identified from statistical tables. The significance level from the statistical tables can be directly compared to 100 - 95 = 5 percent.

If the comparison of average scale scores across two years was made for only one of the five groups, there would be a significant difference between the average scale scores for the two years at a significance level of less than 5 percent. However, because we are interested in the difference in average scale scores across the two years for all five of the groups, comparing each of the significance levels to 5 percent is not adequate. Groups of students defined by shared characteristics, such as racial/ethnic groups, are treated as sets or families when making comparisons. However, comparisons of average scale scores for each pair of years were treated separately, so the steps described in this example would be replicated for the comparison of other current and previous year average scale scores.

Using the FDR procedure to take into account that all comparisons are of interest to us, the percents of confidence in the example are ordered from largest to smallest: 62, 35, 20, 4, and 1. In the FDR procedure, 62 percent confidence for the group 4 comparison would be compared to 5 percent; 35 percent for the group 5 comparison would be compared to 0.05 x (5-1)/5=0.04=4 percent;[4] 20 percent for the group 1 comparison would be compared to 0.05 x (5-2)/5=0.03=3 percent; 4 percent for the group 3 comparison would be compared to 0.05 x (5-3)/5=0.02=2 percent; and 1 percent for the group 2 comparison (actually slightly smaller than 1 prior to rounding) would be compared to 0.05 x (5-4)/5=0.01=1 percent. The procedure stops with the first contrast found to be significant. The last of these comparisons is the only one for which the percent confidence is smaller than the FDR procedure value. The difference between the current year's and previous years' average scale scores for the group 2 students is significant; for all of the other groups, average scale scores for current and previous year are not significantly different from one another. In practice, a very small number of counterintuitive results occur when the FDR procedures are used to examine between-year differences in subgroup results by jurisdiction. In those cases, results were not included in this report.

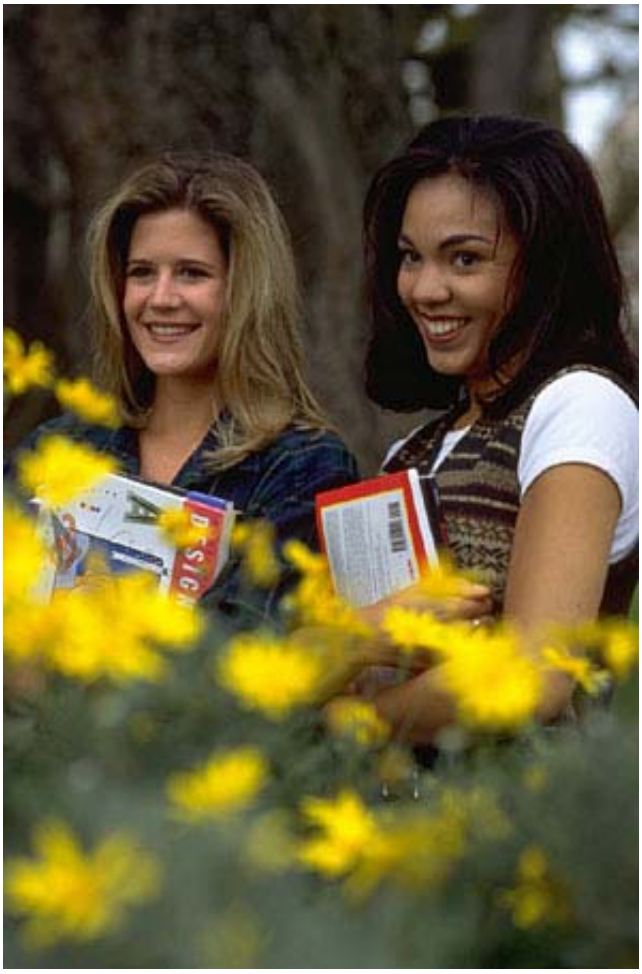
---

1. Miller, R. G. (1981). *Simultaneous Statistical Inference* (2nd ed.). New York: Springer-Verlag.

2. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society,* Series B, no. 1, 289–300.

3. Williams, V. S. L., Jones, L. V., and Tukey, J. W. (1999). Controlling Error in Multiple Comparisons with Examples From State-to-State Differences in Educational Achievement. *Journal of Educational and Behavioral Statistics, 24*(1), 42–69.

4. The level of confidence times the number of comparisons minus one divided by the number of comparisons, or 0.05 x (5-1)/5 = 0.04 = 4 percent.



Understanding NAEP Reporting Groups

NAEP results are provided for groups of students defined by shared characteristics–gender, race or ethnicity, school's type of location, and eligibility for free/reduced-price school lunch. Based on participation rate criteria, results are reported for subpopulations only when sufficient numbers of students and adequate school representation are present. The minimum requirement is at least 62 students in a particular subgroup from at least five primary sampling units (PSUs).[1] However, the data for all students, regardless of whether their subgroup was reported separately, were included in computing overall results. Definitions of the subpopulations are presented below.

Gender: Results are reported separately for males and females.

Race/Ethnicity: In all NAEP assessments, data about student race/ethnicity is collected from two sources: school records and student self-reports. Prior to 2002, NAEP used students' self-reported race as the primary race/ethnicity reporting variable. Starting in 2002, the race/ethnicity variable presented in NAEP reports is based on the race reported by the school. When school-recorded information is missing, student-reported data are used to determine race/ethnicity. The mutually exclusive racial/ethnic categories were White, Black, Hispanic, Asian/Pacific Islander, American Indian (including Alaska Native), and Other. Information based on student self-reported race/ethnicity is available on the NAEP Data Tool (http://nces.ed.gov/nationsreportcard/naepdata/).

Type of Location: Results from the 2003 assessment are reported for students attending schools in three mutually exclusive location types: central city, urban fringe/large town, and rural/ small town.

Central city: Following standard definitions established by the Federal Office of Management and Budget, the U.S. Census Bureau (see http://www.census.gov/) defines "central city" as the largest city of a Metropolitan Statistical Area (MSA) or a Consolidated Metropolitan Statistical Area (CMSA). Typically, an MSA contains a city with a population of at least 50,000 and includes its adjacent areas. An MSA becomes a CMSA if it meets the requirements to qualify as a metropolitan statistical area, has a population of 1,000,000 or more, its component parts are recognized as primary metropolitan statistical areas, and local opinion favors the designation. In the NCES Common Core of Data (CCD) locale codes are assigned to schools. For the definition of central city used in this report, two locale codes of the survey are combined. The definition of each school's type of location is determined by the size of the place where the school is located and whether or not it is in an MSA or CMSA. School locale codes are assigned by the U.S. Census Bureau (see http://www.census.gov/). For the definition of central city NAEP reporting uses data from two CCD locale codes: large city (a central city of an MSA or CMSA with the city having a population greater than or equal to 25,000) and midsize city (a central city of an MSA or CMSA having a population less than 25,000). Central city is a geographical term and is not synonymous with "inner city."

Urban fringe/large town: The urban fringe category includes any incorporated place, census designated place, or non-place territory within a CMSA or MSA of a large or midsized city and defined as urban by the U.S. Census Bureau, but which does not qualify as central city. A large town is defined as a place outside a CMSA or MSA with a population greater than or equal to 25,000.

Rural/small town: Rural includes all places and areas with populations of less than 2,500 that are classified as rural by the U.S. Census Bureau. A small town is defined as a place outside a CMSA or MSA with a population of less than 25,000, but greater than or equal to 2,500.

Results for each type of location are only compared across years 2000 and after. This is due to new methods used by NCES to identify the type of location assigned to each school in the Common Core of Data (CCD). The new methods were put into place by NCES in order to improve the quality of the assignments, and they take into account more information about the exact physical location of the school. The variable was revised in NAEP beginning with the 2000 assessments.

Eligibility for Free/Reduced-Price School Lunch: As part of the Department of Agriculture's National School Lunch Program, schools can receive cash subsidies and donated commodities in turn for offering free or reduced-price lunches to eligible children. Based on available school records, students were classified as either currently eligible for free/reduced-price school lunch or not eligible. Eligibility for the program is determined by students' family income in relation to the federally established poverty level. Free lunch qualification is set at 130 percent of the poverty level, and reduced-price lunch qualification is set at 170 percent of the poverty level. Additional information on eligibility may be found at the Department of Agriculture web site (http://www.fns.usda.gov/cnd/lunch/). The classification applies only to the school year when the assessment was administered (i.e., the 2002–03 school year) and is not based on eligibility in previous years. If school records were not available, the student was classified as "Information not available." If the school did not participate in the program, all students in that school were classified as "Information not available."

---

1. For the national NAEP assessments prior to 2002, a PSU is a selected geographic region (a county, group of counties, or metropolitan statistical area). Since 2002, the first-stage sampling units are schools (public and nonpublic) in the selection of the combined sample.

Where to Find More Information

The NAEP Reading Assessment
The latest news about the NAEP 2003 reading assessment and the national results can be found on the NAEP web site at http://nces.ed.gov/nationsreportcard/reading/results/. The individual snapshot reports for each participating state and other jurisdictions are also available in the state results section of the web site at http://nces.ed.gov/nationsreportcard/states/. *The Nation's Report Card: Reading Highlights 2003* may be ordered or downloaded from the NAEP web site. *The Nation's Report Card: Reading 2003* will be available at the NAEP web site in 2004. *The Reading Framework for the 2003 National Assessment of Educational Progress,* on which this assessment is based, is available at the Internet address http://www.nagb.org/pubs/read_fw_03.pdf.

Additional Results from the Reading Assessment
For more findings from the 2003 reading assessments, refer to the NAEP 2003 results at http://nces.ed.gov/nationsreportcard/naepdata/. The interactive database at this site includes student, teacher, and school variables for all participating states and other jurisdictions, the nation, and the four regions. Data tables are also available for each jurisdiction, with all background questions cross-tabulated with the major demographic variables.

Technical Documentation
For explanations of NAEP survey procedures see Allen, N. L., Donoghue, J. R., and Schoeps, T. L. (2001). *The NAEP 1998 Technical Report*. (NCES 2001–509). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Technical information may also be found on the NAEP web site (http://nces.ed.gov/nationsreportcard/reading/results2003/interpret-results.asp).

Publications on the inclusion of students with disabilities and limited-English-proficient students
Olson, J. F., and Goldstein, A. A. (1997). *The Inclusion of Students with Disabilities and Limited English Proficient Students in Large-Scale Assessments: A Summary of Recent Progress* (NCES 97–482). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Mazzeo, J., Carlson, J. E., Voelkl, K. E., and Lutkus, A. D. (2000). *Increasing the Participation of Special-Needs Students in NAEP: A Report on 1998 Research Activities* (NCES 2000–473). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Lutkus, A. D., and Mazzeo, J. (2003). *Including Special-Needs Students in the NAEP 1998 Reading Assessment, Part I: Comparison of Overall Results With and Without Accommodations* (NCES 2003–467). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Lutkus, A. D. (forthcoming). *Including Special-Needs Students in the NAEP 1998 Reading Assessment, Part II: Results for Students with Disabilities and Limited English Proficient Students* (NCES 2003–468). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

To Order Publications
Recent NAEP publications related to reading are listed on the reading page of the NAEP web site and are available electronically. Publications can also be ordered from:

Education Publications Center (ED Pubs)
U.S. Department of Education
P.O. Box 1398
Jessup, MD 20794–1398

Call toll free: 1-877-4ED PUBS (1-877-433-7827)
TTY/TDD: 1-877-576-7734
FAX: 1-301-470-1244

The NAEP State Report Generator was developed for the NAEP 2003 reports by Phillip Leung, Jilei Yin, Julian Rosse, Paul Gazzillo, Mike Narcowich, Nancy Mead, Anthony Lutkus, Forton Wimbush, Arlene Weiner, and Patricia Hamill.

What is The Nation's Report Card?

THE NATION'S REPORT CARD, the National Assessment of Educational Progress (NAEP), is a nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, history, geography, and other fields. By making objective information on student performance available to policymakers at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement is collected under this program. NAEP guarantees the privacy of individual students and their families.

NAEP is a congressionally mandated project of the National Center for Education Statistics, within the Institute of Education Sciences of the U.S. Department of Education. The Commissioner of Education Statistics is responsible, by law, for carrying out the NAEP project through competitive awards to qualified organizations.

In 1988, Congress established the National Assessment Governing Board (NAGB) to oversee and set policy for NAEP. The Board is responsible for: selecting the subject areas to be assessed; setting appropriate student achievement levels; developing assessment objectives and test specifications; developing a process for the review of the assessment; designing the assessment methodology; developing guidelines for reporting and disseminating NAEP results; developing standards and procedures for interstate, regional, and national comparisons; determining the appropriateness of all assessment items and ensuring the assessment items are free from bias and are secular, neutral, and nonideological; taking actions to improve the form, content, use, and reporting of results of the National Assessment; and planning and executing the initial public release of National Assessment of Educational Progress reports.

The National Assessment Governing Board

Darvin M. Winick, Chair
President
Winick & Associates
Dickinson, Texas

Honorable Jo Ann Pottorff, Vice Chair
State Legislator
Topeka, Kansas

Amanda P. Avallone
Assistant Principal and
Eighth-Grade Teacher
Summit Middle School
Boulder, Colorado

Barbara Byrd-Bennett
Chief Executive Officer
Cleveland Municipal School District
Cleveland, OH

Carl A. Cohn
Clinical Professor
Rossier School of Education
University of Southern California
Los Angeles, CA

Shirley V. Dickson
Program Director, Literacy
Education Commission of the States
Denver, CO

Edward Donley
Former Chairman
Air Products & Chemicals, Inc.
Allentown, Pennsylvania

John Q. Easton
Executive Director
Consortium on Chicago School Research
Chicago, IL

Honorable Dwight Evans
State Legislator
Philadelphia, Pennsylvania

Sheila M. Ford
Principal
Horace Mann Elementary School
Washington, DC

David W. Gordon
Superintendent of Schools
Elk Grove Unified School District
Elk Grove, CA

Catherine (Katy) M. Harvey
Principal
Bethesda-Chevy Chase High School
Bethesda, Maryland

Juanita H. Haugen
Local School Board Member
Pleasanton, California

Honorable Dirk A. Kempthorne
Governor of Idaho
Boise, Idaho

Kathi M. King
Twelfth-Grade Teacher
Messalonskee High School
Oakland, ME

Kim Kozbial-Hess
Fourth-Grade Teacher
Fall-Meyer Elementary School
Toledo, Ohio

Honorable Ronnie Musgrove
Governor of Mississippi
Jackson, Mississippi

Diane Ravitch
Senior Research Scholar
New York University
New York, New York

Mark D. Reckase
Professor
Measurement and Quantitative Methods
Michigan State University
East Lansing, MI

Sister Lourdes M. Sheehan, R.S.M.
Associate General Secretary
United States Catholic Conference
Washington, DC

John H. Stevens
Executive Director
Texas Business and Education Coalition
Austin, Texas

Honorable Michael E. Ward
State Superintendent of Public Instruction
North Carolina Public Schools
Raleigh, North Carolina

Eileen L. Weiser
Member, State Board of Education
Michigan Department of Education
Lansing, MI

Dennie Palmer Wolf
Director, Opportunity & Accountability Initiatives
Annenberg Institute for School Reform
Providence, RI

Honorable Grover (Russ) Whitehurst
(Ex-officio Member)
Director
Institute of Education Sciences
U.S. Department of Education
Washington, DC

Charles E. Smith
Executive Director, NAGB
Washington, DC